

AD _____

CONTRACT NUMBER DAMD17-96-C-6061

TITLE: 3-D Ultrasound Vascularity Assessment for Breast Cancer Diagnosis

PRINCIPAL INVESTIGATOR: Paul L. Carson, Ph.D.

CONTRACTING ORGANIZATION: University of Michigan
Ann Arbor, Michigan 48109-1274

REPORT DATE: September 1997

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

DTIC QUALITY INSPECTED 2

19980310 041

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED	
	September 1997	Annual (1 Sep 96 - 31 Aug 97)	
4. TITLE AND SUBTITLE		5. FUNDING NUMBERS	
3-D Ultrasound Vascularity Assessment for Breast Cancer Diagnosis		DAMD17-96-C-6061	
6. AUTHOR(S)			
Paul L. Carson, Ph.D.			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)		8. PERFORMING ORGANIZATION REPORT NUMBER	
University of Michigan Ann Arbor, Michigan 48109-1274			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
Commander U.S. Army Medical Research and Materiel Command Fort Detrick, Frederick, Maryland 21702-5012			
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION / AVAILABILITY STATEMENT		12b. DISTRIBUTION CODE	
Approved for public release; distribution unlimited			
13. ABSTRACT (Maximum 200)			
<p>This project is to improve the diagnosis and management of patients with breast cancer through development and evaluation of 3D ultrasound imaging and quantification techniques emphasizing vascularity. Progress this first year included some development of our existing methods for 3D breast ultrasound. These developments included refinements of power mode and frequency shift Doppler acquisition, quantification and display of vascularity. Most effort was directed to three longer-range topics: 1) image registration using the mutual information cost function for evaluation/demonstration of image compounding and sequential examinations; 2) approaches to real time image-based registration (IBaR) for 3D slice positioning; 3) the effort to image lactic ducts in 3D. Outstanding results were obtained both in the automated image registration of scans in one exam using two different color flow imaging modes and in scans separated by 45 days during therapy. Initial tests of IBaR were repeated and refined; image processing was programmed for automated recognition of trackable regions of the images and concepts were developed for application to real-time ultrasound systems. New scans were performed for lactic duct delineation and software developed for rotating through the planes containing a given duct.</p>			
14. SUBJECT TERMS		15. NUMBER OF PAGES	
Breast Cancer, 3D and Ultrasound		103	
		16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT		18. SECURITY CLASSIFICATION OF THIS PAGE	
Unclassified		Unclassified	
19. SECURITY CLASSIFICATION OF ABSTRACT		20. LIMITATION OF ABSTRACT	
Unclassified		Unlimited	

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

X Where copyrighted material is quoted, permission has been obtained to use such material.

X Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

X Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

 In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

X For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

 In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

 In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

 In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.


Paul L. Larson
PI - Signature

9/22/97
Date

I. TABLE OF CONTENTS

FRONT COVER

STANDARD FORM (SF) 298

FOREWORD

I. TABLE OF CONTENTS.....	1
II. INTRODUCTION.....	2
IV. BODY.....	3
A. Clinical Trials and Vascularity Measures.....	3
1. Validation/improvement of vascularity measures.....	3
2. Clinical Trials of ~30 subjects	3
3. Trials data analysis and write-up	3
B. Compound and Sequential Imaging.....	3
1. Develop automated image registration by mutual information	3
2. Process ultrasound images to evaluate/demonstrate compound and sequential* imaging therewith.....	3
C. Image-based Registration of Slice Position [IBaR].....	4
1. Real time speckle decorrelation for image plane positioning.....	4
D. Digital ductography development.....	4
1. Ductographic analysis of selected clinical trial cases	4
E. Figures.....	5
V. CONCLUSIONS.....	8
Clinical Trials	8
Compound and Sequential Imaging.....	8
Image-based Slice Positioning [IBaR].....	9
Digital ductography development.....	9
VI. REFERENCES	9
VII. APPENDICES	10

1. The 3D and 2D Color Flow Display of Breast.....	11
2. Cumulative Power Distribution Function: Technique for Defining a Stable Intravascular Point for Normalizing Fractional Moving Blood Volume Estimates with Power Doppler US.....	24
3. Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline warped geometric deformations	63
4. 3D Doppler Image Signal Quantification of Breast Masses	75
5. Power Doppler Ultrasound findings of blood vessels in a case of breast cancer: Comparison made in 2D and 3D.....	94

II. INTRODUCTION

The research to be conducted over the four year period of this contract centers around three areas 1) expansion of clinical trials capabilities performed under an ongoing NIH grant and their continuation after those trials are completed; 2) image registration using the mutual information cost function for evaluation/demonstration of image compounding and sequential examinations; 3) approaches to real time image-based slice positioning; 4) the, as yet, unique effort to image lactic ducts in 3D and perform preliminary evaluation of their diagnostic potential. Most of the research will be performed in parallel by various investigators. The following time line for the first two years was given in the funded scope of work.

Table 1. Time Line for Present and Next Project Periods	Year 1	Year 2
Clinical Trials		
Validation/improvement of vascularity measures		++
Clinical Trials of ~30 subjects		++
Trials data analysis and write-up	--	
Compound and Sequential Imaging		
Develop automated image registration by mutual information	++++	+ - + -
Process ultrasound images to evaluate/demonstrate compound and sequential* imaging therewith	++	+
Image-based Slice Positioning [IBaR]		
Real time speckle decorrelation for image plane positioning	+ -	- - - -
Digital ductography development	++	++
Ductographic analysis of selected clinical trial cases		

*Auto registration of breast images from multiple exams + = concentrated effort - = modest effort

The following series of deliverables and associated materials are included in this annual progress report. Included are a listing of papers, abstracts and manuscripts, published and submitted.

- Item 1: First generation prototype software descriptions for image registration.
- Item 2: Descriptions of IBaR computations and speculation on methods of real time IBaR.
- Item 3: Report on first generation prototype software for visual identification of ducts.

IV. BODY

Progress has been made in the four areas of emphasis of this grant as listed below.

A. Clinical Trials and Vascularity Measures

Although minimal work was scheduled in this area for the first year, opportunities for progress has arisen and several advances have been made beyond those anticipated.

1. Validation/improvement of vascularity measures

Reference 3 (Appendix 3), describes a new technique for determining the signal for 100% blood that has been developed in part under this contract for improving and automating the normalization of Doppler measures for approximating fractional blood volume (NPD) and perfusion ($v \cdot NPD$). As defined in Reference/Appendix 3, NPD is the normalized, power-weighted pixel density and $v \cdot NPD$ is the mean detectable blood velocity times NPD. Over the next year, we hope to begin implementing this measurement technique for testing in 3D data sets.

2. Clinical Trials of ~30 subjects

Scheduled for the last 18 months, we have done two patients to test new 3D acquisition techniques including cardiac gating for capturing the higher color flow signal during systole without signal dropout or averaging over the cardiac cycle. This used a new, breast-dedicated high end ultrasound system and new software allowing storage of only desired images into the system's cine-loop memory and direct image data storage to our 3D workstations. 3D acquisition and display of vascular trees, especially small vessels, appears to be improved significantly. Fig. 1 shows an example of a 3D power mode image of an elongated infiltrating ductal carcinoma, obtained with a new hand-held scanner and direct image recording (without a frame grabber), but prior to the high speed image acquisition and cardiac gating. The images are much more interpretable with the vascularity shown in color.

3. Trials data analysis and write-up

Analysis of previous trials was conducted in part under this grant, resulting in a publication and two submissions, References/Appendices 1, 4 and 5.

B. Compound and Sequential Imaging

The most exciting advances with the greatest long term significance involve the automatic registration of two or more 3D image sets to add to the information and quality of the resulting set or to demonstrate changes between two sets.

1. Develop automated image registration by mutual information

Software originally developed for fusion of images from multiple modalities has been refined to run efficiently with a minimum of operator interaction. This software and some of its earlier applications are explained in Reference/Appendix 3. This paper is our best available prototype software description for image registration.

2. Process ultrasound images to evaluate/demonstrate compound and sequential* imaging therewith

The first attempts at registering 3D ultrasound image sets with each other have been outstanding successes. We have assumed all along that the ultrasound images would require processing to smooth the coherent speckle and thus reduce noise resulting from the algorithms trying to register speckle spots that should be independent in two different image sets. Surprisingly, the ultrasound image sets tested reveal enough mutual information over and above the speckle noise to allow excellent registrations under conditions of significant tissue strain between images. One example shows the results of registering a 3D frequency-shift color Doppler image (f-CDI) volume of an infiltrating ductal carcinoma with a 3D power mode color Doppler image (p-CDI) volume obtained a few minutes after the f-CDI scan. Fig. 2 displays

the registration of two sequential exams obtained 45 days apart during the course of chemotherapy. In both examples, the degree of displacement and distortion between the registered sets was surprising, making the outstanding registrations all the more remarkable.

C. Image-based Registration of Slice Position [IBaR]

In our continuing efforts to advance the 3D slice plane positioning, we have developed a number of software tools which allow analysis of data collected using our hand-held position encoding system. The software developed under MATLAB takes advantage of existing subroutines to ensure greater portability and should allow transfer into 3D display routines in the AVS Express software being used for 3D display of the breast vasculature. In its present form, the software assumes a Gaussian beam profile and pure speckle regions with Rayleigh statistics. Therefore, the B-scan intensity correlation function will also be Gaussian and the decorrelation rate is proportional to step size for uniform scanning over a stationary medium. The basic technique is to compute the average autocorrelation function for a specific region of interest over a limited number of sequential scan frames. This function is curve fit to a Gaussian, and based on a calibrated, depth-dependent, "apparent" elevational beamwidth, the step size is determined from the fitted variance. Key issues include ROI size and number of frames for accurate step size estimation, the validity of actual Rayleigh statistics, and the influence of noise on the Gaussian fitting algorithm. In addition, speckle region identification has been automated using first order statistics and will soon be tested on breast data collected in the clinical studies.

In initial tests of the new software, a number of image sets from linear, uniform scanning of phantoms at different step sizes were collected and analyzed for the Diasonic Spectra using the 10 MHz linear array. The 10 MHz probe was selected to provide a smaller elevational beam width than we have tested in the past and is closer to that presently being used for breast scanning (see conclusions). Fig. 3a shows the correlation curves computed for 10 frames and averaged over a square centimeter. The estimated step size, Fig. 3b, is within 5% of the actual frame spacing. These results are very similar to those recently published (Reference 6) which indicate that the technique as implemented in the new software is working well even with the reduced elevational beam width of the 10 MHz probe. We are presently obtaining data on the General Electric LOGIQ 700 scanner using the new M12 micro-slice probe. This probe is used to acquire the 3D image sets for the breast vasculature study and has electronic focusing in the elevational (slice thickness) dimension. This focusing will provide a greater depth of field for the ultrasound image which should make the decorrelation in the elevational direction more stable and thus improve prospects for success with IBaR. However, while elevational focusing improves image quality, the smaller beamwidth of this probe may pose some challenges for IBaR since the decorrelation will occur much more rapidly.

1. Real time speckle decorrelation for image plane positioning

This topic is discussed below under conclusions.

D. Digital ductography development

Using the anatomy of the breast, software tools were developed to help the user manually find, map and review ductal structures in the a volume of B mode ultrasound data. Initially, the user identifies the relative location of the nipple to the data volume. The user then searches the volume by rotating a plane about the nipple. When a duct-like structure is seen, the user draws a line along the duct. The software then enables him to rotate the search plane around this line to verify if it is a duct, find the portions of it not in the initial search plane, and identify visible lobules feeding it. During this process, the user can mark any feature as part of the ductal tree, ultimately creating a 3D mapping of ductal structures within the volume.

1. Ductographic analysis of selected clinical trial cases

Three cases have been studied; one is shown in Fig. 4. The large, gray pattern in 3D outlines the observed ducts in this ultrasound field of view. Results should be even more impressive when using the new 2D array (multislice) scan head which we now use in our clinical trials.

E. Figures

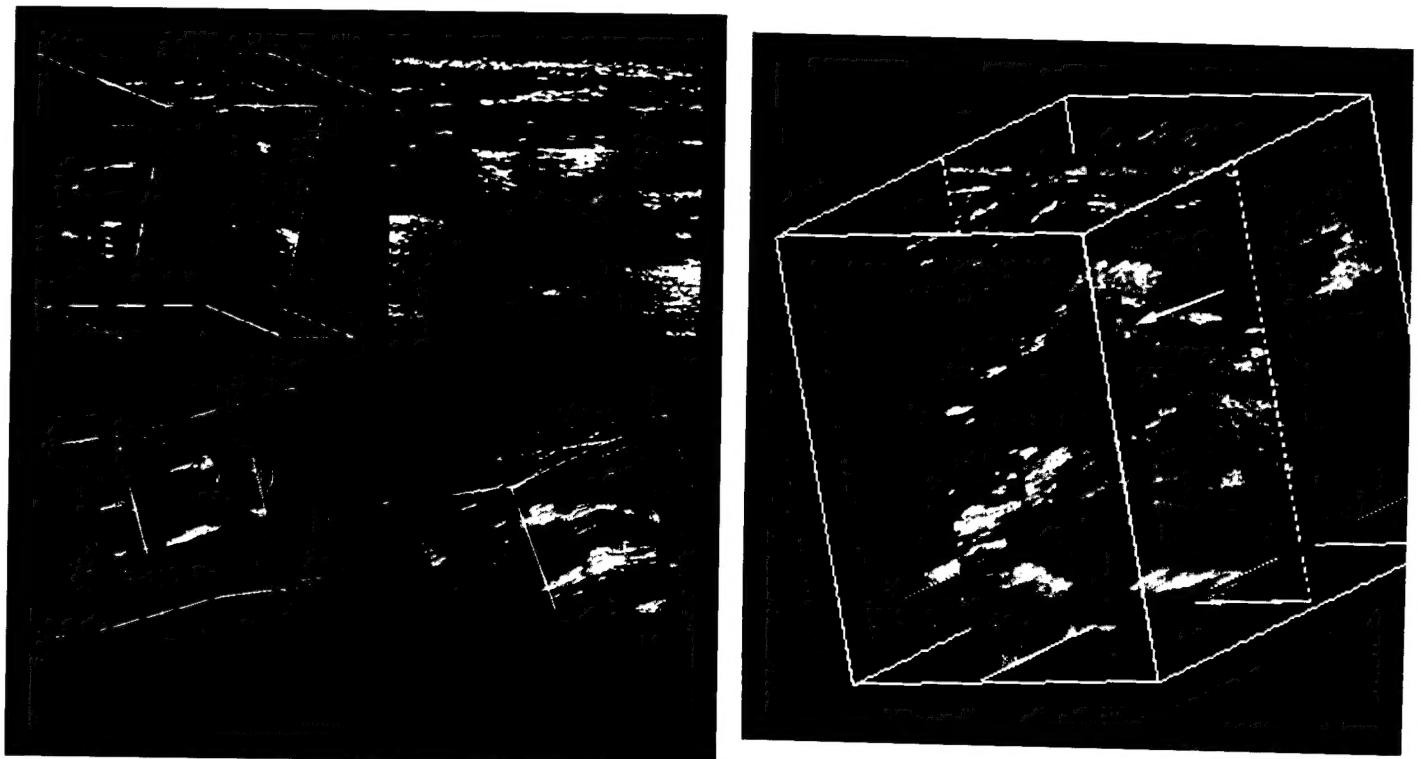


Figure 1. 2-D and 3-D power mode display of an elongated infiltrating ductal carcinoma, shown clockwise from top left in the first print. These gray scale images, taken from color prints to fit this report, do lose some of the information content described. (a) 3-D view demonstrates two planes, one of which is the original scan plane and the second an orthogonal plane which depicts the extent of the hypoechoic mass. (b) 2-D image of the original scan plane which demonstrates a vessel penetrating the mass, evidence for the solid nature of this mass. (c) 3-D image with an ellipsoidal ROI for quantification of vascularity. It is shown as a translucent green ellipsoid, to depict the approximate extent of the lesion, with the penetrating vessel seen as yellow. (d) 3-D display with ROI in green, and isosurface display of vascularity in blue. (e) On the second print, two planes are shown through the mass. The modest vascularity throughout the 3D volume is shown as isosurfaces, with the penetrating vessel indicated by the arrow.

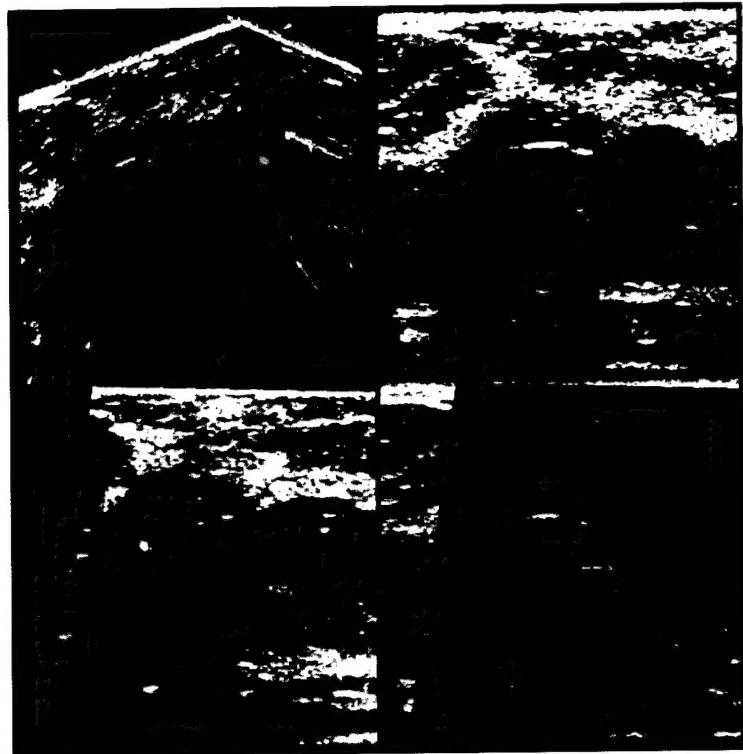


Figure. 2 Clockwise from the top right are results of one of the first two 3D registrations between two ultrasound data sets. The registration is between two sequential exams obtained 45 days apart during the course of chemotherapy for locally advanced infiltrating ductal carcinoma in a 52 year old woman. (a) Three orthogonal slices through the large, hypoechoic, lesion. The two separate data sets, the first set in (b) and the second (transformed) set in (c) did not cover the same part of the breast very well, although that was the intention of the examinations, designed for tracking vascularity and tumor shrinkage or growth during treatment. While shown much better in the color versions of these images, the region of overlap, between the arrows in (a) covered only approximately 1/2 of the scanned volume. The automated 3D registration allows close comparison of vascularity and tumor size and shape change in the same locations of the tumor, particularly in the color displays of these images. Even here, the smoothing of the speckle, more uniform representation of the bright echo regions and attendant increase in contrast between the hypoechoic tumor and surrounding tissues is apparent. There is not a large amount of vascularity in these slices; it is shown as white blobs in these images. In this case there was minimal change between these second and third image sets during the treatment. The tumor almost disappeared by the fourth imaging. A subtraction image (d) highlights differences in the two data sets. If image warping is employed to obtain local spatial match of the structures, the change in the tumor size and shape might well be indicated by the distortion tensor between the two data sets over the region of the registered mass.

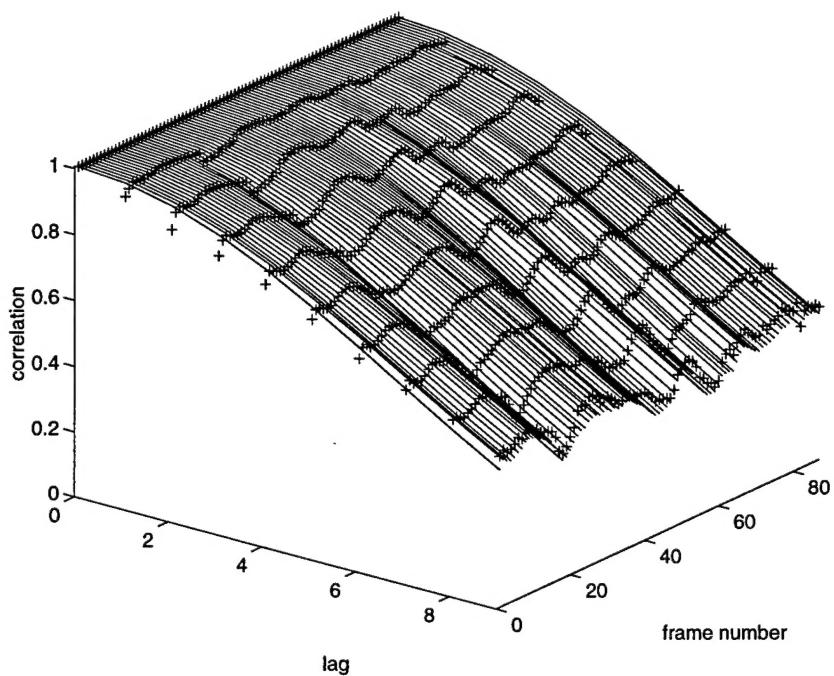


Figure 3a Correlation functions for uniform scanning of a tissue-mimicking phantom. Curves were computed over 10 frames, averaged over a square centimeter, and assigned to the first frame in the series.

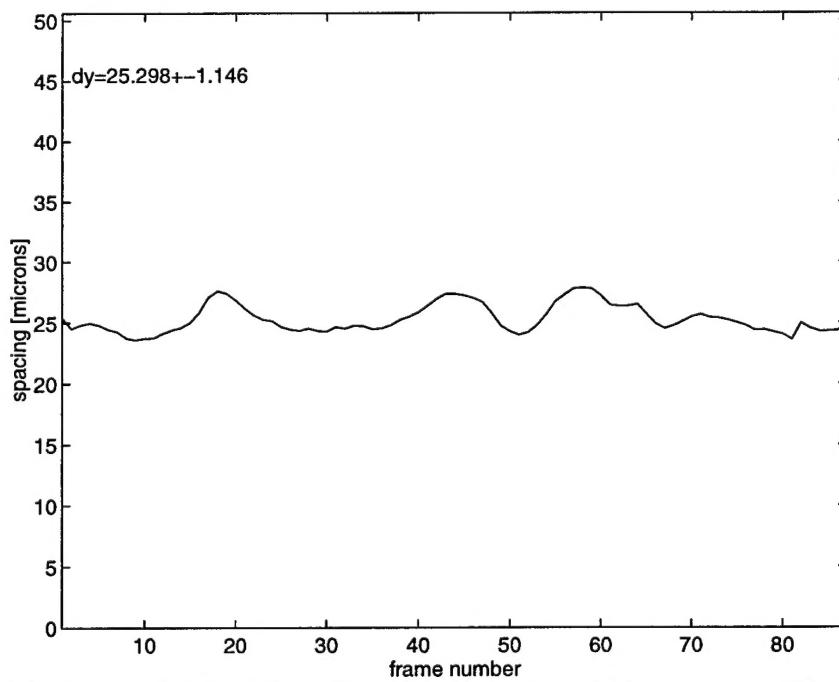


Figure 3b Estimated frame spacing based on Gaussian fits to correlation curves. The actual step size was 25.4 microns.

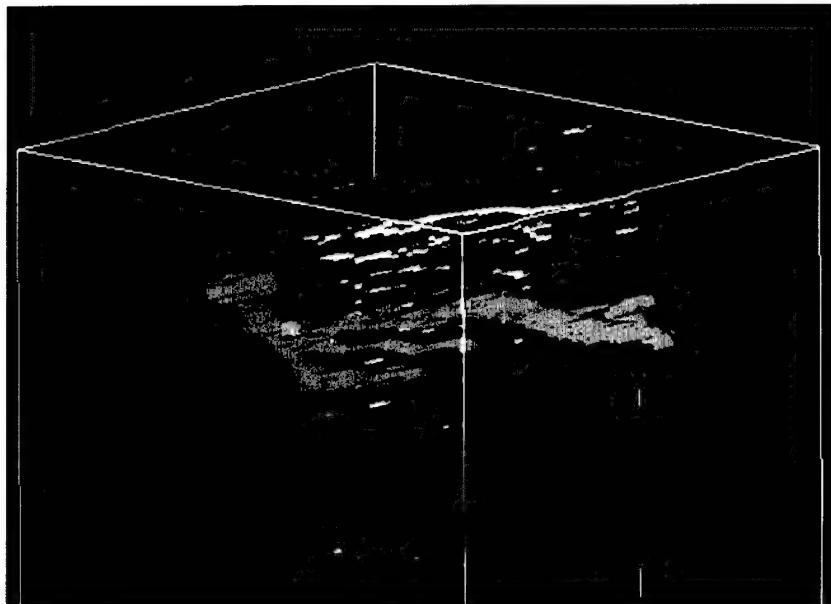


Figure 4. Ductal tree in gray, 3D superimposed on one of the gray scale slices in the volume. This is better seen in color. The Dark layer in the gray scale slice, lying 70% of the way down between the top of the image and the ducts might be thought of as a duct, but careful panning through the slices shows it to be a surface, rather than a tube.

V. CONCLUSIONS

The advances described above should increase the sensitivity and specificity of ultrasound imaging. This will continue to be in conjunction with mammography, but increasingly in conjunction with MRI and in expanded roles in the young and mammographically dense breast.

Clinical Trials

Data analysis and write-up of earlier trials has been completed and contributions were made to advance 3D image acquisition, display and quantification.

Compound and Sequential Imaging

The registration achieved so far is not at pixel level yet to allow accurate pixel-by-pixel correlation of velocity and power information in the calculation of v•NPD, but the registration is much better than can be done with visual identification of corresponding regions of interest and more complete delineation of gray scale and flow structures is provided. Fig. 2 reveals the potential value, respectively, of using the 3D registration for combining information from several types of ultrasound scans to provide an image volume

of improved information content and for following changes in the patient in serial exams. Both, but particularly the latter can revolutionize medical imaging and interpretation. The former should get good enough to allow compound scanning (imaging the same tissues from different viewing angles) in at least some cases and combining dedicated high resolution, (multifocal) gray scale images with f-CDI and/or p-CDI scans. These combinations will reduce speckle and electronic noise, while revealing more of the structure seen only at a few of the image look directions. The combinations also will allow time for higher resolution imaging in each mode while combining the different information from each. The types of serial exams that might be followed include tracking of tumor size and vascularity for early indications of response to treatment, as well as serial follow-up or even screening exams to reveal abnormal growth.

Image-based Slice Positioning [IBaR]

While the results to date are impressive, the possibility of practical implementation also seems very promising. To accurately estimate elevational movement, the frame rates must be high enough such that consecutive images are not completely decorrelated. An alternative scheme is to compute the decorrelation from a limited number of intermediate pulse firings. Interestingly, the pulses transmitted for the color flow imaging may be just the information required because of the high pulse repetition frequency (PRF) for the Doppler estimate and the multiple firings necessary for averaging. Therefore, real-time data processing for IBaR may be possible on information that is already being acquired for another purpose. Scanhead motion would not be constrained unnecessarily in velocity by the low frame rate during Doppler imaging. Alternatively, control of the scanhead firing during multifocal B-mode operation would allow similar processing of data that continues real-time imaging with frame rates at those currently employed. These possible imaging schemes may be explored with the assistance of one of the ultrasound system manufacturers as it will require data processing of the incoming data stream and/or control of the scanhead firing.

Digital ductography development

The functional part of the breast is the glandular tissue which is found in 15-20 lobes which radiate out from the nipple. Each lobe is drained by its central duct which exits from the nipple. Almost all breast cancer occurs in the glandular tissue, and a significant proportion is intraductal. Therefore, being able to find and map the ducts might greatly increase the efficacy of B mode ultrasound for breast cancer screening and diagnosis. Display of the ductal tree in 3D, while panning through the gray scale images should be a significant aid to orientation and anatomy. At the least it will facilitate registration of serial volume image sets of the breast.

VI. REFERENCES

1. Carson PL, Moskalik AP, Govil A, Roubidoux MA, Fowlkes JB, Normolle D, Adler DD, Rubin JM, Helvie M: The 3D and 2D Color Flow Display of Breast Masses, *Ultrasound Med. Biol.*, 23/6, 837-849, 1997.
2. Rubin JM, Bude RO, Fowlkes JB, Spratt RS, Carson PL, Adler RS, Cumulative Power Distribution Function: Technique for Defining a Stable Intravascular Point for Normalizing Fractional Moving Blood Volume Estimates with Power Doppler US, *Radiology*, accepted with minor revisions.
3. Meyer CR, Boes JL, Kim B, Bland PH, Wahl RL, et al., Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline warped geometric deformations. *Medical Image Analysis* 1(3), 195-206, 1997.

4. Carson PL, Fowlkes JB, Roubidoux MA, Moskalik AP, Normolle D, LeCarpentier G, Nattakom S, Rubin JM: 3D Doppler Image Signal Quantification of Breast Masses, Ultrasound Med. Biol., submitted.
5. Nattakom ST, Le Carpenter G, Roubidoux MA, Carson PL, Power Doppler Ultrasound findings of blood vessels in a case of breast cancer: Comparison made in 2D and 3D, Ultrasound Med. Biol., submitted.
6. Chen J-F, Fowlkes JB, **Carson PL**, Rubin JM: Determination of Scan-Plane Motion Using Speckle Decorrelation. Theoretical Consideration and Initial Test, Internat. J. Imaging Systems and Technology, 8, 38-44, 1997.

VII. APPENDICES

1. Carson PL, Moskalik AP, Govil A, Roubidoux MA, Fowlkes JB, Normolle D, Adler DD, Rubin JM, Helvie M: The 3D and 2D Color Flow Display of Breast Masses, Ultrasound Med. Biol., 23/6, 837-849, 1997.
2. Rubin JM, Bude RO, Fowlkes JB, Spratt RS, Carson PL, Adler RS, Cumulative Power Distribution Function: Technique for Defining a Stable Intravascular Point for Normalizing Fractional Moving Blood Volume Estimates with Power Doppler US, Radiology, accepted with minor revisions.
3. Meyer CR, Boes JL, Kim B, Bland PH, Wahl RL, et al., Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin plate spline warped geometric deformations. Medical Image Analysis 1(3), 195-206, 1997.
4. Carson PL, Fowlkes JB, Roubidoux MA, Moskalik AP, Normolle D, LeCarpentier G, Nattakom S, Rubin JM: 3D Doppler Image Signal Quantification of Breast Masses, Ultrasound Med. Biol., submitted.
5. Nattakom ST, Le Carpenter G, Roubidoux MA, Carson PL, Power Doppler Ultrasound findings of blood vessels in a case of breast cancer: Comparison made in 2D and 3D, Ultrasound Med. Biol., submitted.



●Original Contribution

THE 3D AND 2D COLOR FLOW DISPLAY OF BREAST MASSES

PAUL L. CARSON,[†] AARON P. MOSKALIK,[†] ANURAG GOVIL,[†]
 MARILYN A. ROUBIDOUX,[†] J. BRIAN FOWLKES,[†] DANIEL NORMOLLE,[†]
 DORIT D. ADLER,[†] JONATHAN M. RUBIN[†] and MARK HELVIE[†]

Departments of [†]Radiology and [‡]Biostatistics, University of Michigan Medical Center,
 Ann Arbor, MI 48109 USA

(Received 19 December 1996; in final form 8 April 1997)

Abstract—A prospective study was performed in 24 women with breast masses on mammography going on to surgical biopsy. 2D and 3D power mode and frequency shift color flow Doppler scanning and display were compared. Vessels were displayed as rotatable color volumes in 3D, superimposed on gray-scale slices. The latter were stepped sequentially through the imaged volume. Radiologists rated the masses in each display (3D, 2D and videotapes) on a scale of 1 to 5 (5 = most suspicious) for each of six conventional gray-scale and six new vascular criteria. Thirteen masses proved to be benign and 11 were malignant. 3D provided a stronger subjective appreciation of vascular morphology and allowed somewhat better ultrasound discrimination of malignant masses than did the 2D images or videotapes (specificities of 85%, 79% and 71%, respectively, at a sensitivity of 90%). Only in 3D did the vascularity measures display a trend towards significance in this small study. © 1997 World Federation for Ultrasound in Medicine & Biology.

Key Words: Blood, Angiogenesis, Blood volume, Perfusion, Ultrasound, Doppler, Image processing, Three-dimensional imaging, Breast cancer, Diagnostic imaging.

INTRODUCTION

In order to further standardize breast-lesion vascularity measures, allow better appreciation of the 3D vascular morphology, and evaluate information provided in the new power-mode color-flow images, 3D acquisition and display techniques were developed. A study was performed using 3D ultrasound examinations of women with breast masses scheduled for subsequent breast biopsy. Each examination consisted of 1 or more power-mode 3D scans, a frequency-shift 3D scan of the mass and surrounding tissue, (usually) a video tape record of these scans, and 2D films of interesting gray-scale and vascular features at the time of the examination.

Angiogenesis associated with tumor growth has been claimed to be an indicator of malignancy and possible predictor of metastasis (Folkman and Shing 1992; Weidner et al. 1992; Weidner 1995). Studies suggest that there are fundamental differences between benign and malignant vasculature. As in many other

malignancies, rapid, intense and/or extended enhancement with CT, radionuclide and MRI contrast agents are a feature of all or part of most malignant breast lesions (Sakki 1974; Chang et al. 1982; Watt et al. 1986; Wahl et al. 1991; Chenevert et al. 1994).

In recent years, therefore, there has been increased interest in using Doppler ultrasound when examining solid breast lesions to differentiate malignant from benign lesions. Some earlier studies began with the simple criterion of the presence of detectable vasculature near or in the lesion as a positive indicator of malignancy (Dixon et al. 1992; Adler et al. 1990; Cosgrove and Lees 1994; Carson et al. 1992; Lagalla et al. 1994). Other indicators were added, with varied levels of success. The robustness of the simplest criterion, presence of detectable vasculature near or in the lesion, as well as most other vascularity measures, suffers because the threshold is dependent on machine sensitivity. Measures based on color velocity values appear to be more robust than color pixel density (Fein et al. 1995).

Other studies have utilized the velocity information provided by spectral Doppler and color flow Doppler to characterize malignancy by calculating parameters, such as peak flow velocity, resistive index,

Address correspondence to: Paul Carson, Ph.D., Professor of Radiology, Kresge III, Rm. R2315, University of Michigan Medical Center, Ann Arbor, MI 48109-0553 USA. E-mail: pcarson@umich.edu

and pulsatility index (McNicholas et al. 1993; Kuijpers et al. 1994; Madjar et al. 1994; Peters-Engl et al. 1995). The results from these studies have been highly mixed, as might be expected for local measures of complex tumor structures using incomplete search procedures.

Another method to detect malignancy involves observation of the vascular pattern. This can entail counting vessels (Heilenkotter and Jagella 1993; McNicholas et al. 1993; Madjar et al. 1994), counting vessels that penetrate a mass (Bergonzi et al. 1993; Calliada et al. 1994; De Albertis et al. 1995), and evaluating vessel morphology such as tortuosity (Heilenkotter and Jagella 1993; Calliada et al. 1994). The excellent quality and relative simplicity of displaying color-flow image data in 3D was illustrated by Picot et al. (1993). 3D volumetric quantitative measures, such as color pixel density and mean color value, have also been investigated (Pretorius et al. 1992; Carson et al. 1993; Huber et al. 1994; Kedar et al. 1995). Normalization of the signal levels to the level from the largest blood signal at the same depth, and other 3D measures of the power mode signal have been introduced (Carson et al. 1993). Most of these vascularity studies yielded similar results, with a quoted sensitivity of around 90% and a specificity of around 70%, or *vice versa*, depending on how the criteria were chosen. Clearly, there are often differences between malignant and benign vascularity that can be detected by Doppler ultrasound, but the value of Doppler images in addition to pulse echo studies is still debated strongly.

The difficulty in providing accurate differentiation between benign and malignant masses includes more than just finding definitive or dependable measures. The vascularity in some patients may be changed systemically or locally around lesions due to other factors, such as the increased vascularity resulting from needle aspiration (Kedar and Cosgrove 1994). Systemic acoustic differences between patients can be accounted for by comparison to the contralateral breast (Scherzinger et al. 1989; Blohmer et al. 1995). Other factors that can cause ambiguity are differences between machines, machine settings, depth of the lesion, properties of overlying tissues, and Doppler angle. In most studies, many of these factors are ignored, making the results difficult to reproduce.

Diagnosis does not have to rely on vascularity independent of other indicators. Some investigators with careful technique and good equipment have become very skilled at cancer discrimination using pulse-echo gray-scale analysis (Venta et al. 1994; Leucht et al. 1988). When the objective was clearly defined as prevention of unnecessary biopsies and the gray-scale

criteria for designating a lesion as benign were very stringently set. Stavros et al. (1995) achieved a false-negative rate of only 0.5% in a population of 750 biopsy-proven cancers. In that study, decisions based on the ultrasound appearance would have eliminated 420 of the 750 scheduled biopsies, with 2 malignant masses being missed. If independent measures such as vascularity were added only to distinguish benign or probably benign from malignant lesions, this extremely small false-negative rate might be achievable in more centers.

A recent large multicenter study with modern color-flow systems (Goldberg et al. 1993) suggested that ultrasound evaluation of breast lesions has been significantly improved in the past few years. Although some of the participating investigators concluded that the additional information from color-flow was marginal, others reported that the improvement in specificity of ultrasound over mammography was statistically significant only if Doppler results were added to the pulse-echo results. The addition of color-flow Doppler to the B-mode improved the diagnostic accuracy by an amount approximately equal to the benefit of adding ultrasound to mammography (Cosgrove and Lees 1994; Cosgrove et al. 1993). Positive predictive value almost doubled over mammography alone and the false-negative rate was only 1.7%.

The possibility of new methods for regional vascularity assessment (Dymling et al. 1991; Carson et al. 1993) helped motivate the development of medical color-flow imaging with display of the Doppler signal power, instead of the estimated Doppler frequency shift (Rubin et al. 1994). The combination of high-resolution power Doppler and frequency-shift color Doppler imaging now provides the opportunity to assess the accuracy of regional vascularity (Rubin et al. 1995) and provide more qualitative studies with available instrumentation, such as the research reported here.

METHODS

Women were selected after identification of a mass based on mammography and/or clinical examination. Masses of very low suspicion (*e.g.*, most fibroadenomas) were not included. All patients in the study were scheduled for surgical biopsy of the mass as part of their clinical management. Informed consent was obtained from all participating women. 3D scans were acquired using the system described and depicted in Moskalik et al. (1995). The system consisted of a conventional ultrasound scanner (Spectra VST, Diasonics, Inc., San Jose, CA, USA), a modified mammography unit that applies mild compression to immo-

bilize the breast, a frame grabber that is used to acquire video images from the ultrasound imager, and a motorized transducer positioner mounted on the mammography unit. A typical scan through an acoustically transparent window produces a 3D volume of ultrasound data of the suspicious region of the breast from a series of approximately coronal 2D images. These are separated typically by approximately $3/f$ (mm), where f is the Doppler imaging frequency in MHz, or, preferably, by half the 6 dB focal slice thickness. These empirical relations are based on our experience with maintaining enough coherence between adjacent slices to allow continuous volume rendering of the vasculature. At the 6-MHz Doppler frequency of our 6–10 MHz linear array, the step size was 0.5 mm. This is accomplished by mounting the imaging array to the positioner. The breast is placed in the mammographic unit, and the array is moved in 0.5-mm steps across the region of interest as the frame grabber acquires the 2D images. Data from the 50–120 images are then transferred to a workstation. With the linear array scan-head employed for this study, the images were typically 38 mm by 40 or 50 mm. Power-mode Doppler ultrasound was performed at a PRF of 700Hz, and an ensemble length of 14, the minimum sample length and the maximum gain not producing frequent color noise. The 2D images utilized for comparison with the 3D included the best free-hand images photographed during a conventional color-flow study, which also served to help identify the mass region for the 3D scan. Also available to the readers when desired for the 2D interpretations were the individual digital images recorded during the slowly stepped slices from the 3D acquisitions. This helped assure that at least as much ultrasound information was potentially available for the 2D as for the 3D readings. The videotapes usually included much of the preliminary searching for the correct region for 3D scanning, and often included the real-time display of the slowly stepped 3D scanning.

Two types of analysis are being performed with the resulting data. The analysis reported in this paper is comparative assessment of visualization using 3D and 2D imaging techniques, with and without vascularity criteria, as described below. The other analysis currently in progress is measurement of the signals produced in the 2 color-flow modes quantitatively to estimate vascularity features in various 3D regions of interest in and around the suspected masses (Carson *et al.* 1995).

Of 37 attempted studies, 24 were performed satisfactorily and were reviewed by 2 of the 3 reviewers in both the 3D display and 2D images. Eleven masses were pathologically proven to be cancer, five were fibroadenomas, and eight were other benign masses.

Videotapes were available or technically acceptable on 14 of 24 cases, allowing videotape analysis on five cancerous, three fibroadenomatous, and six other benign masses. The order of viewing of the 3 display types by an individual reader was rotated randomly. In the conduct of research where humans were the subjects, the investigators adhered to the policies regarding the protection of human subjects as prescribed by 45 CFR 46 and 32 CFR 219 (Protection of Human Subjects).

There are 2 dominant methods of displaying relatively simple 3D objects, such as a few blood vessels, which fill a small fraction of the imaged volume. One method commonly employed in MR angiography is to trace rays in straight lines through the 3D data set and make a projection image like a radiograph with the dominant tissue representing x-ray attenuation as the detected blood flow. Usually, only the pixel with the strongest power-mode signal or highest frequency shift is represented along a given ray. This method is also being employed in several commercial and commercial-prototype packages for 3D ultrasound displays. As employed in our work, the second method requires setting a threshold Doppler power-mode signal level or threshold Doppler frequency shift, depending on the imaging mode. All adjacent pixels at that signal level are connected by 3D surfaces, termed isosurfaces.

With either display method, the selected 3D data can be viewed from any given angle. A rotating 3D display of vasculature that does not fill the volume too densely is effective at revealing the 3D morphology. With substantially more computing power, the 3D display can be rotated and shown in correct positional relationship with selected 2D slices of the full gray-scale/color-flow data. Addition of the gray-scale image or full gray-scale plus color image requires rapid texture rendering, as well as the isosurface rendering. An example of the Doppler isosurface-plus-gray-scale type of display is shown in Fig. 1. Throughout this study, the 3D display was offered as any desired combination of: 1. Stepping through the color flow images in their correct orientations; 2. Continuous 3D rendering of the vasculature, while either (a) stepping through the stack of 2D images, (b) rotating the vascular volume along with 1 chosen gray-scale slice, or (c) rotating the vasculature only.

Recently, we have developed 2 methods of improving the display technique with respect to the time required fully to appreciate the 3D vasculature and its relation to the gray-scale features of the lesion. These have been demonstrated on data from a few of the subjects in this study. The acquired stack of gray-scale slices is viewed sequentially with no vascular isosurface. In one of the new methods, a translucent render-

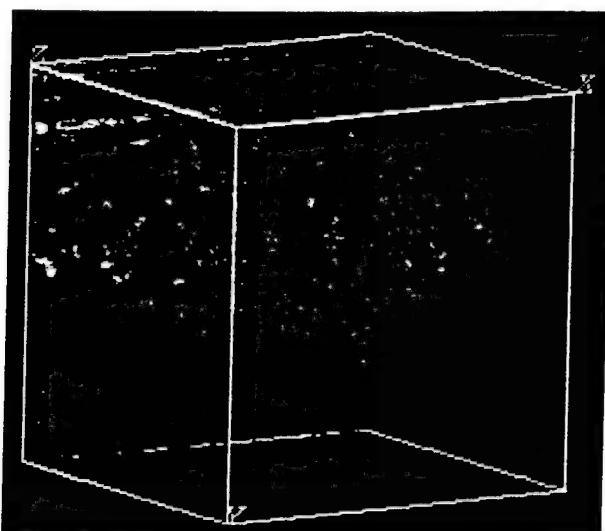


Fig. 1. 3D vascularity display utilizing isosurface rendering of power-mode data taken from a 45-y-old subject with invasive intraductal carcinoma. The gray-scale slice plane is shown in correct spatial relationship to the vessels. The last gray-scale plane is shown here so as not to obscure any of the vascularity. The position of the gray-scale information is usually moved through the volume-rendered vascularity. The entire volume is then rotated further to increase geometrical perception of the vascularity in relation to 1 or more gray-scale slices.

ing of the vasculature is shown superimposed on the gray-scale stack after every 3 to 4 steps. In the other method, the 3D vascular rendering is shown in a second window beside the sequential display of the gray-scale slices. A frame outline or a nearly transparent gray-scale slice is stepped through the 3D vascular rendering to reveal the location in 3D of the gray-scale slices being displayed in the first window. Both methods make it easier to correlate gray-scale information with vasculature, without hiding gray-scale information by the vascular rendering. When very fast display update is desired, a preprocessed stack of images is shown, although this is now becoming less necessary. The rendering of the images into a video or computer-displayed movie is also automated using a command line interface to our 3D rendering and image-processing software (Advanced Visualization System, AVS).

After imaging, the masses of the 24 women were then classified fibroadenoma, other benign or cancer; the first 2 categories were then grouped for analysis as noncancerous. With the videotape display, not all the masses were evaluated. The utility of the gray-scale and vascular criteria in discriminating malignant from benign masses was evaluated. To make this evaluation, assumptions concerning the relative frequency of malignant and benign masses in the population of

interest and of the relative costs of misclassification were required. It was assumed that 20% of all masses are malignant, and 80% of all masses are benign, consistent with a typical mix of positive and negative findings at biopsy in a screening situation. There is a tremendous potential flexibility in the ratio of true-positive and true-negative findings in mammography clinics. The fact that typically only 20% of the cases recommended for biopsy prove malignant is a reflection of a broad medical judgment that the cost, in a very general sense of the word, is 5 times higher for a false-negative than for a false-positive. This ratio was assumed for the discriminant analysis.

With these assumptions, the linear discriminant using both the gray-scale and vascularity indices that minimized the expected total generalized cost was estimated. Probabilities of misclassification were estimated using cross-validated resubstitution, that is, each observation in the sample (the training sample) was withdrawn, the discriminating function reestimated, and then the withdrawn observation was classified. This technique produces a relatively conservative estimate of the error rate.

RESULTS

A gray-scale example of the 3D rendering of the vasculature (usually displayed in color) is shown in Fig. 1. This example represents a 3D ultrasound scan of a mass in a 45-y-old woman with microcalcifications in the left breast detected at mammographic screening. A nonpalpable 7-mm intraductal carcinoma (comedo type) with negative margins was diagnosed in a general area of florid fibrocystic change. A more complete illustration of the 3D display can be viewed on the Internet. The Internet images are compressed with some loss of resolution for improved display speed and storage considerations. The address on the World Wide Web is: http://www.med.umich.edu/ultrasound/breast_anim/develop/Case38.html. Instructions are given for viewing this animation with shareware such as Apple's Sparkle. In the Internet example, one of the new display techniques is illustrated. The gray-scale images are stepped through the volume in their relative 3D positions at a fixed rate. Following every 4 ultrasound slices, a view of the 3D Doppler signal isosurfaces is shown in place of the gray-scale for the duration of four image cycles.

The images were obtained in a coronal plane, with the central ultrasound beam at a 3 o'clock angle, but shifted approximately 2.5 cm cephalad to the level of the nipple. The beginning and ending frames, in the front and back of Fig. 1, respectively, are approximately 50 mm and 31 mm from the nipple. The small

Table 1. Six gray-scale criteria were defined with scales from 1 to 5, 5 being those characterized as the most malignant.

	Criterion description	Benign extreme (1)	Malignant extreme (5)
1	Margin of mass/smoothness	Smooth, no lobulations	Microlobulations
2	Shape	Spherical	Highly irregular
3	Sound attenuation of mass	Enhancement	Strong shadowing
4	Echogenicity of mass	Cyst or echogenic mass	Hypoechoic
5	Orientation of mass	Flattened	Taller than wide
6	Visibility of margins	Sharply defined	Diffuse margins

Scaling of Criterion 4, texture of mass, is not a linear scale because it is used to first distinguish between cystic and solid masses. Therefore, a rating of 1 corresponds to anechoic (cystic) and ratings 2-5 are solid masses, with 2 being most echogenic and 5 being hypoechoic.

cluster of microcalcifications seen on the mammogram corresponds to the upper left of slice 40, proximal to its acoustic shadow. This is at the end of a probable earlier surgical scar, apparent from a sharp shadow seen at approximately that location in the preceding few images. Significant vascularity wrapping around the margins of a tissue region is not seen until slices 109 to 133. The tissue partially enclosed by vascularity is slightly hypoechoic, but otherwise unremarkable.

To help standardize assessment of the gray-scale and 3D vascularity measures in breast lesions, the criteria in Tables 1 and 2 were developed for this study. These criteria were taken from the literature and our own experience. The central vascularity of tumors greater than 2 cm diameter was not evaluated, specifically because it was expected that the degree of necrosis would be highly variable and not diagnostic. Central vascularity is being evaluated in a companion quantitative study of these subjects (Carson *et al.* 1995). Table 3 shows results on several of the vascular criteria that individually were considered good discriminators between benign and malignant masses, because they received higher ratings on the 1 to 5 scale for cancer than for benign masses. For example, mean ratings from the 3D display for criterion 1, vascularity in the outer 1 cm of the mass, was 3.3 for malignant masses and 2.2 for fibroadenomas and other benign masses. With all 3 displays, vascularity from the mass border

to 1 cm outside the mass (criterion 2), provided less discrimination between cancer and fibroadenomas than did vascularity in the mass. This is consistent with the diagnostic criterion that fibroadenomas and cancer have extensive external vascularity, but the vessels rarely penetrate the fibroadenoma (Bergonzi *et al.* 1993; Calliada *et al.* 1994; De Albertis *et al.* 1995). The last 2 vascular criteria in Table 2 were positively correlated with malignancy, but were not as discriminative individually. In the analysis below, they did not contribute positively to discrimination using criteria averages and were considered no further.

The radiologists reported a subjective improvement in understanding of the overall morphology and distribution of vessels resulting from the 3D images, as indicated by the relatively good benign/malignant discrimination in 3D by criteria 1-3. There was also an impression by the readers that hand searches in 2D for the correct angles to reveal shunting might be done as well, or better, in some cases, than in 3D. Overall, the morphology and extent of vascularity were believed to be seen much better in power mode than in frequency-shift mode.

In comparison with the 2D displays, there was thought to be little subjective advantage in reading gray-scale criteria with the 3D renderings. What improvement was observed was from seeing the gray-scale slices sequentially in their correct 3D locations.

Table 2. Eight vascular criteria were defined, also scaled from 1 to 5, 5 being the rating of the most malignant character of a criterion.

	Criterion description	Benign extreme (1)	Malignant extreme (5)
1	Vascularity in mass (outer cm)	None	Extensive (> 25% of mass)
2	Vascularity outside mass (to 1 cm away)	None	Extensive (> 25% of area)
3	Visible shunt vessels	None	Very clear or more than 2
4	Vessels wrapping around mass	None	Total involvement > 240°
5	Tortuosity of vessels	Gently curving	Very tortuous
6	Apparently related vessels, beyond 1 cm	None	Unusually large, fast flow
7	Unusual vascularity, no associated mass	None	Multiple, large, fast flow
8	General enlargement of breast vessels	None	Extensive

Table 3. Examples of results from four of the vascular criteria.

	3D	2D	Video
1. Mass vascularity, outer 1 cm	-	-	-
Benign	2.2	2.0	2.4
Fibroadenoma	2.1	2.0	2.3
Malignant	3.3	2.9	2.9
2. Vascularity outside mass 1 cm	-	-	-
Benign	2.4	2.2	2.6
Fibroadenoma	3.2	2.7	2.5
Malignant	3.7	2.9	2.8
3. Visible shunt vessels	-	-	-
Benign	2.1	1.5	1.2
Fibroadenoma	1.6	1.6	2.0
Malignant	2.7	2.4	2.6
4. Vessels wrapping around mass	-	-	-
Benign	1.9	1.7	1.6
Fibroadenoma	2.8	2.2	2.5
Malignant	3.5	2.6	2.4

The numbers represent the mean ranking for the given criterion over all readings of that mass type: malignant ($n = 11$), fibroadenomas ($n = 5$) and other benign masses ($n = 8$). Noting that a rating of 1 is most benign and a 5 most malignant in appearance, 3D generally received higher rankings because more vascularity was detected. 3D also provided a wider separation between malignant and other masses on all but Criterion 3.

The results of the discriminant analysis described in Methods are given in Table 4 and plotted as the solid lines in Figs. 2–4. Videotape is probably inferior to 2D and 3D, although the small sample size may make the comparison misleading. 3D appears slightly better than 2D, but the discriminator lines are similar.

Relative contributions of gray-scale and vascular indices in the linear discrimination were evaluated for each of the 3 displays. The p -values for significance in the discriminators are given in Table 5. Only in the 3D display did vascularity display a trend towards significance in this small study.

ROC curves for decision rules using gray-scale alone and gray-scale plus vascularity are shown for the 3D display in Fig. 5a. In Fig. 5b, ROC curves for decision rules using gray scale plus vascularity are drawn for the 3 different displays. The accuracies are the areas under these curves (Swets and Pickett 1982),

with 3D having the largest accuracy. The discrimination threshold and numbers in Table 4 can be changed from those arrived at by the assumed cost factors. Selecting a 90% ratio of true-positives to all cancers (sensitivity in the group studied), the specificities for 3D, 2D and videotape displays were 85, 79 and 71%, respectively. At that same sensitivity of 90%, the specificities for 3D displays with gray-scale criteria and with gray-scale plus vascularity were 85 and 81%, respectively.

The visual readings as a function of lesion diameter are revealing. Lesion diameters from the largest among the mammography, ultrasound and pathology estimates are given. The diameters in mm as means (and ranges) are: malignant, 16 (6–25); fibroadenoma, 17 (14–23); other benign masses 18, (7–29). A plot of mean 3D gray-scale rating as a function of this maximum diameter shows strong discrimination of the smaller, 6–17 mm diameter, cancers from the benign masses of equivalent sizes. This discrimination is particularly good for cancer and fibroadenomas. 3D vascularity readings separate most of the largest, 15–25 mm, cancers (Fig. 6a). For fibroadenomas, the figure suggests an increase in vascularity with lesion diameter. A combination of the 3D vascularity plus three times the 3D gray-scale reading, all divided by four, reflects the relative weighting (the slope) of the discriminator shown in Fig. 2. This combination variable provides good separation of malignant from nonmalignant masses, as indicated in Fig. 6b. There is some suggestion that addition of lesion size information may aid the discrimination provided by the gray-scale and vascularity criteria shown in Fig. 2.

DISCUSSION AND CONCLUSIONS

New techniques for acquiring, displaying and interpreting 3D Doppler power-mode and frequency-shift color-flow and gray-scale images of the breast were developed and employed in this study. Vascularity, particularly the vascular morphology and extent, were considered subjectively to be revealed better in 3D than in the 2D displays. This was true even though

Table 4. Results using the gray scale and vascularity linear discriminators.

Condition Classified	Benign		Cancer		Specificity TN/(TN + FP)	Pos Pr Val TP/(TP + FP)	Neg Pr Val TN/(TN + FN)
	Benign TN	Cancer FN	Benign FP	Cancer TP			
Videotape	7	2	2	3	77.8%	60.0%	77.8%
2D	10	3	1	10	76.9%	76.9%	90.9%
3D	10	3	0	11	76.9%	78.6%	100.0%

It was assumed 20% of all masses in the population to be studied are malignant and the financial and personal costs of a false-negative classification are 5 times the cost of a false positive.

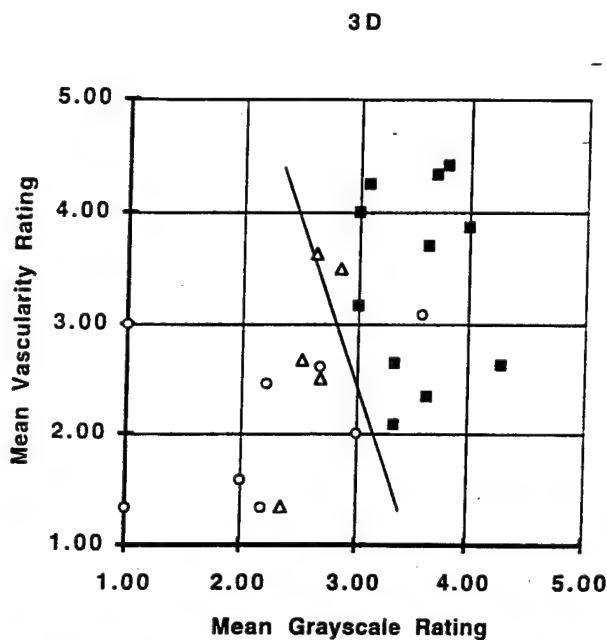


Fig. 2. Scatter plot from the 3D readings of the average ratings of the six vascular criteria versus the average ratings of the six selected gray scale criteria. Benign masses are shown as open symbols, fibroadenomas as (Δ) and other benign masses as (\circ). The malignant masses are shown as (\blacksquare), clustered near the upper right-hand corner, as expected.

The linear discriminator is shown as the solid line.

all the slices of the 3D image acquisition were available for viewing in the 2D assessment, as desired. The frequency-shift Doppler should emphasize the rapid flow associated with many tumors. Vascular morphology was better seen in power mode than in frequency-shift mode, but the 2 Doppler modes were not compared here. More extensive and angle-independent display of breast vascularity from compounded power and frequency-shift mode color-flow imaging has been illustrated by Moskalik *et al.* (1995), using reregistration of the different views. However, that technique was not practical for these clinical trials.

For statistical analysis, the incidence assumption that 20% and 80% of the masses are malignant and nonmalignant, respectively, were based on a reasonable mix for the group of patients usually going to biopsy, rather than the group with mammographic or palpable masses in this study. As listed in Tables 3 and 5, as well as Fig. 5b, videotape was probably inferior to 2D and 3D in the discrimination of cancer (lower sensitivity, specificity and negative predictive values). 3D was slightly better than 2D, with a higher sensitivity and negative predictive value.

Only in 3D did vascularity display a trend towards significance. Statistical significance is, of course, a function of sample size. If the distributions had been

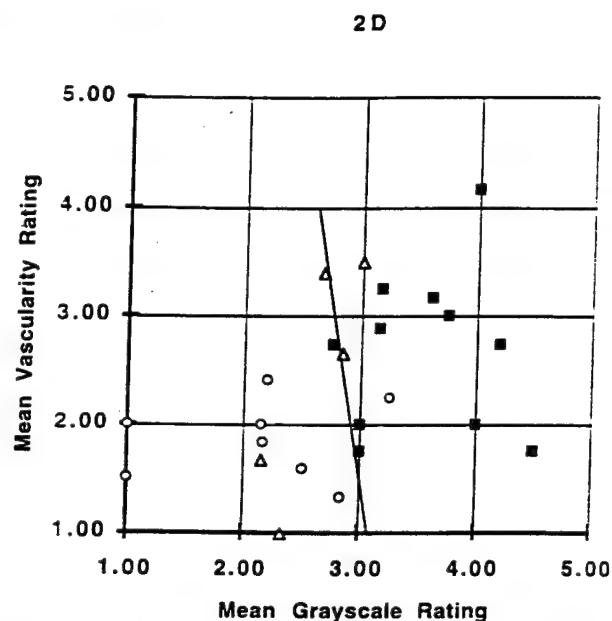


Fig. 3. Mean visual ratings from the 2D images. Vascular and gray scale criteria, the symbols and discriminator line are as in Fig. 2.

the same, but with a sample size 3 times larger, vascularity would have been significant at $p = 0.05$. Had it been 4 times larger and with the same results, vascularity would have been significant at $p = 0.01$. With the discriminators determined with vascularity and gray-

VIDEOTAPE

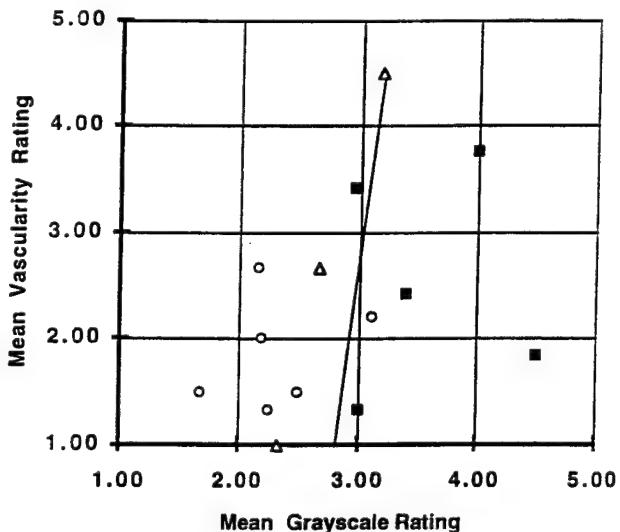


Fig. 4. Mean visual ratings from the videotapes of the actual examinations. Vascular and gray scale criteria: the symbols and discriminator line are as in Fig. 2.

Table 5. The p values for significance in the gray-scale and vascularity discriminators.

	Gray-scale	Vascularity
Videotape	0.0031	0.66
2D	0.0001	0.57
3D	0.0001	0.19

scale measures, the sensitivity, specificity, positive predictive value and negative predictive value of 3D were 100%, 77%, 79% and 100%, respectively. For these numbers to be meaningful for reducing unnecessary biopsies, a modest positive predictive value must be achieved while maintaining a very high negative predictive value in a much larger study, including a mix of all cases going to biopsy, not just those with identifiable masses. This might be feasible in a large study by increasing the assumed relative cost of false-negatives in determining the discrimination line in a new version of Fig. 2. It could then be determined if, with these newer techniques, the results are better than those (sensitivity, specificity, positive predictive value and negative predictive value of 98.4%, 68%, 38%, 99.5%) of the large gray-scale study of Stavros et al. (1995), or the as yet not fully published multicenter study with gray-scale and color-flow imaging (Goldberg et al. 1993, Cosgrove and Lees 1994). It might also prove in the long run that vascularity imaging in the dense breast would reveal some suspected cancers that are not mammographically detected and might otherwise be followed only by periodic exams.

Pulse echo ratings were quite similar for at least the 3D displays and 2D images. This is as expected, because, unlike the vascular display, we have not undertaken the more difficult task of surface rendering of gray-scale structures to show them in 3D much better than in 2D. It was noted subjectively that, even without the 3D vascular rendering, it was helpful to see the gray-scale images in their proper relative locations in space, at least as a substitute for personally performing the gray-scale scanning. Subjective impressions were that the power mode revealed vascular patterns more fully than did the frequency-shift mode of color-flow images. However, that distinction between modes was not the purpose of the study and was not rated systematically.

The videotape was meant to simulate the physicians performing the examinations themselves or, more accurately, observing the real-time examination on a remote monitor. Because videotaped images are of inferior quality and that technique is rarely employed clinically, examinations were not disqualified when there was both no time for separate video recording

and the scanner was in a condition such that we were unable simultaneously to record the images digitally and on videotape. The image reviewers rarely found additional information on the videotape records and had a clear preference for the other 2 displays.

Because of the rather large fluctuations in any one vascular or gray-scale criterion in the visual analyses, taking the maximum reading from any 1 vascular and 1 gray-scale criterion did not provide

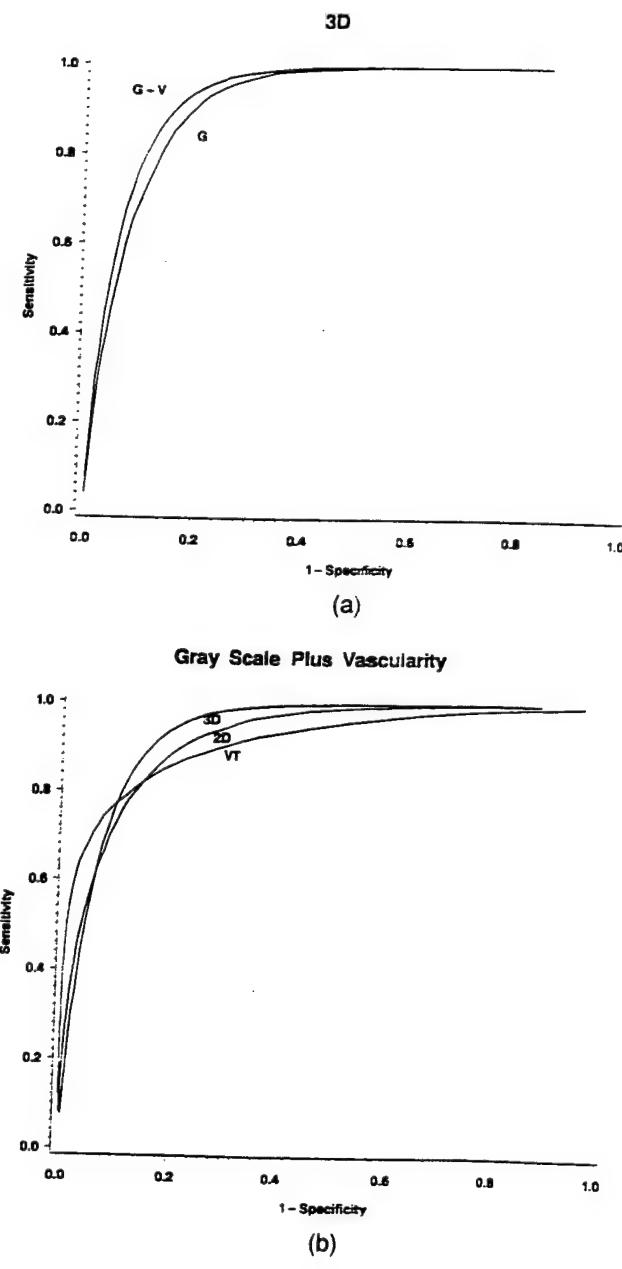


Fig. 5. ROC curves of the data in Figs. 2-4. The curves are for decision rules using: (a) Gray-scale alone and gray-scale plus vascularity for the 3D display, Fig. 2. (b) Gray-scale plus vascularity for the 3 different displays.

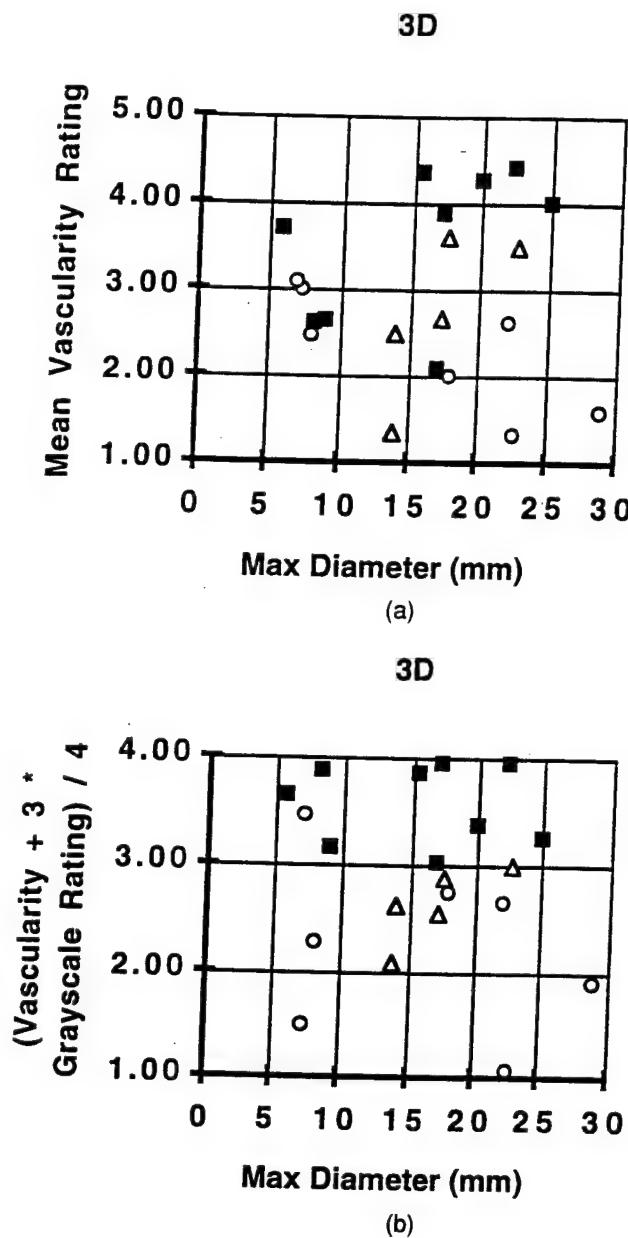


Fig. 6. Visual readings as a function of the largest mass diameter as obtained from 3 different measures of the mass dimensions. Readings of: (a) Mean vascularity, and (b) A combination in which gray-scale readings are weighted by 3 times those of the vascularity readings.

very good discrimination. Other combinations, such as the average reading from the three criteria correlating best with cancer discrimination in the sample, performed only marginally better than any 1 criterion. Only the average readings of the 6 mass-related vascular criteria (excluding criteria 6 and 7 in Table 2) and the 6 gray-scale criteria provided good discrimination. It might be hoped that overall vascu-

larity throughout the 3D volume scanned would be a good discriminator because simple methods might be developed to test that criterion. In this study, however, vascularity throughout the breast was not a particularly good or even average discriminator. Examination time constraints did not allow a consistent assessment of pulsatility or of a high density of small vessel, multidirectional flow in the frequency-shift color-flow images. These might be reassessed, particularly as small-vessel sensitivity of frequency-shift color-flow imaging is increased via the instrumentation or contrast agents.

The apparent increase in vascular penetration of a gray-scale-defined cancer (Fig. 7a, b), compared with a fibroadenoma (Fig. 8a, b) suggests that vascularity in the outer 1-cm radius of the mass, or its ratio to the vascularity immediately outside the mass, should be quite important. However, in this small study, these criteria, 1 and 2, did not stand out over the other mass-related vascular criteria. The observation that the 2 youngest women with breast cancer also had the most vascular lesions is consistent with the fibroadenoma data of Cosgrove *et al.* (1993).

This research was, in part, an effort to develop more specific hypotheses about visual discrimination of breast cancer with 3D imaging of vascularity and gray-scale features. It would be appropriate to test the conclusions and discriminant values of this study on an independent set of subjects. It is hoped that such study could provide more information on whether or not vascularity is a particularly good discriminator, and possibly detector, of cancer in women less than 40 or 45 y old.

At the initiation of this study, the types of 3D displays employed were cumbersome, using affordable equipment. However, 3D hardware and software are improving rapidly and, indeed, are already at or close to functional speed for these techniques for less than \$25000 worth of computer components and software, prior to packaging as a medical product. Now that dedicated chips for 3D display in PCs are beginning to be produced in mass quantities, prices are falling rapidly.

The main issues in evaluating 3D versus 2D techniques with ultrasound are different than those in the other high-resolution tomographic modalities. Conventional 2D ultrasound techniques generally include rapid, intuitive, real-time searches from numerous angles to find any views that best show features of interest through the overlying aberrating tissues. By comparison, any 1 3D scan might not have the best viewing location and angle for most of the region of interest, even when guided by real-time 2D imaging. However, 3D offers systematic scanning and imaging

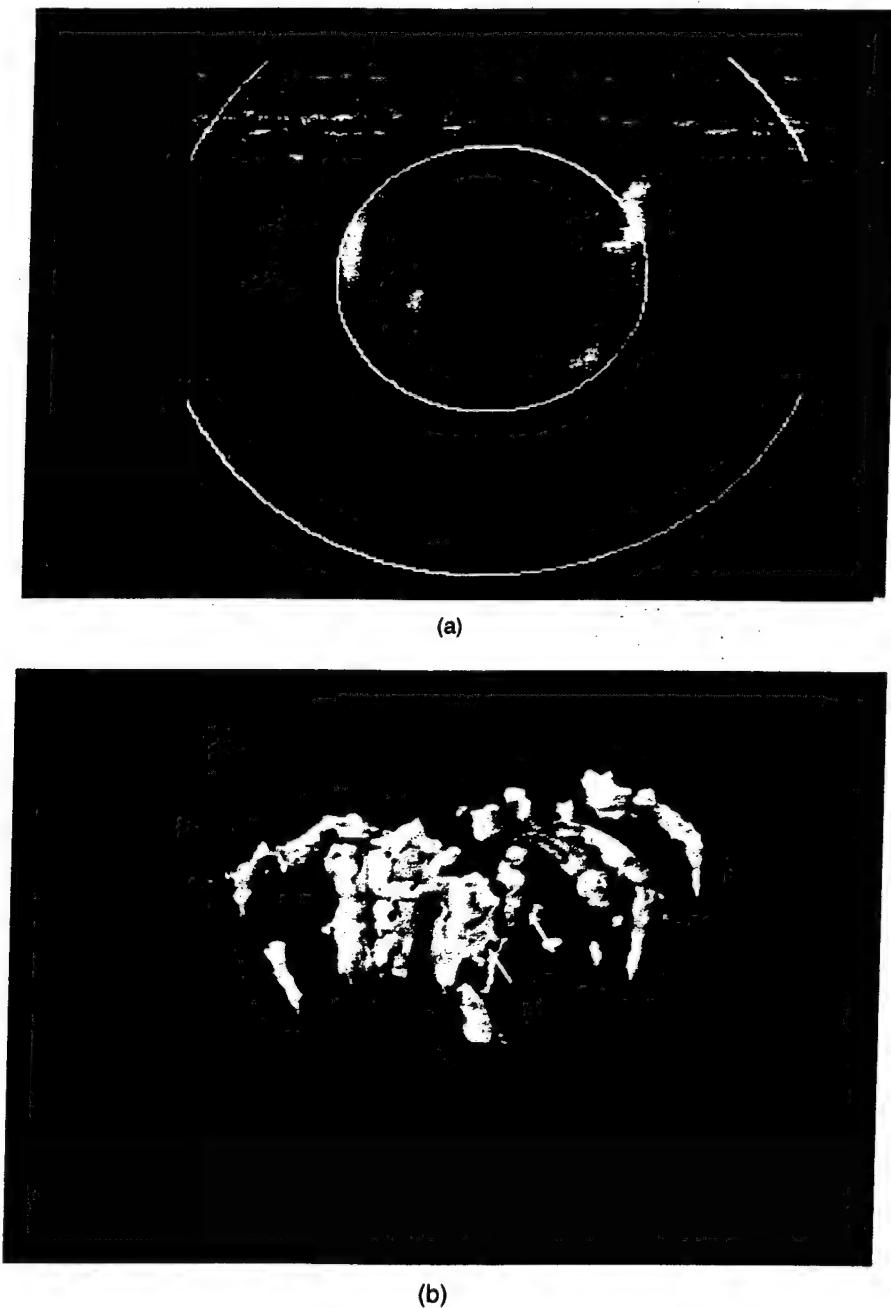


Fig. 7. Power mode-detected vascularity from a 2.5×2 cm malignant nodule in a 33-y-old woman is displayed as bright regions of relatively smooth texture in (a) 2D and (b) 3D. (a) Power-mode color-flow image approximately through the center of the mass, with outlines of the identified mass region and a region 1 cm beyond. (b) Power-mode-detected vascularity displayed as isosurfaces, with a pulse echo slice shown at the edge of the scanned volume on the nipple side. Here, the nodule is outlined at its estimated gray-scale and palpable borders by a translucent ellipsoid at the right posterior edge. Vessels surround and penetrate this ellipsoid. Example penetrating vessels are marked by solid arrows.

of the entire mass region, scanned and sampled at rates optimized for pulse echo or color flow Doppler sensitivity. The 3D display also can be presented at a convenient time with full flexibility. The display offers a

high information content per view and reveals correct spatial relationships even in views that cannot be obtained directly in 2D scans. This paper represents a multidisciplinary development and assessment of state-

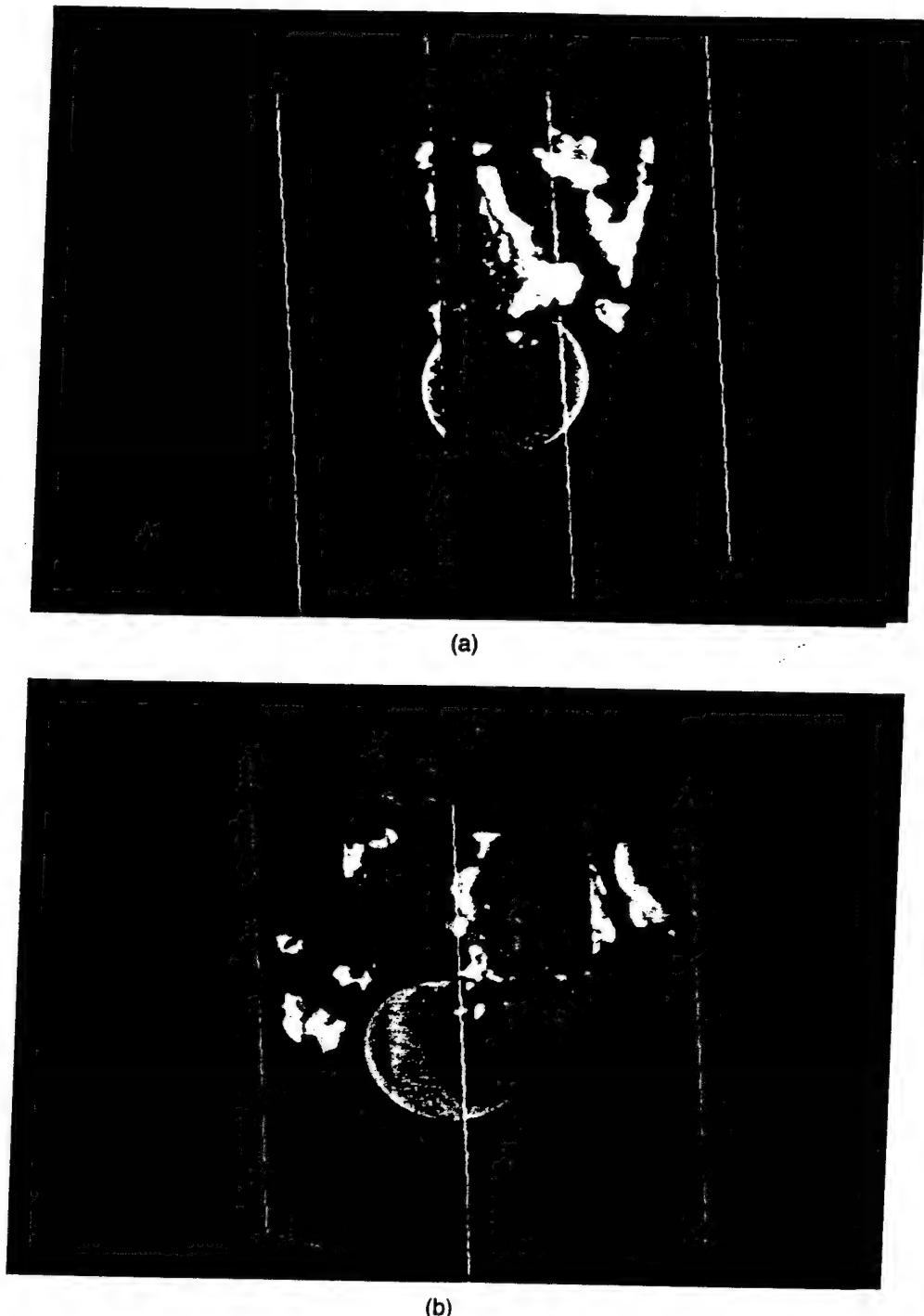


Fig. 8. 3D displays of power mode-detected vascularity around a relatively vascular, 2.3-cm fibroadenoma in a 45-year-old woman. (a) and (b) A gray-scale plane through the fibroadenoma is shown from each side, along with the ellipsoidal volume selected to follow the gray-scale borders of the mass.

of-the-art 3D ultrasound techniques and offers encouraging, but not yet statistically significant, results as to whether or not 3D imaging and display can offer additional information to that provided by conven-

tional real-time ultrasound techniques in the evaluation of breast lesions.

Acknowledgments—This work was supported in part by USPHS

grant IRO1CA55076 from the National Cancer Institute and by the U. S. Army Medical Research and Materiel Command under Contract No. DAMD17-96-C-6061. The views, opinions and/or findings contained herein are those of the authors and should not be construed as an official Department of the Army position, policy or decision. Dr. Mihra Taljanovic contributed significantly in performance and interpretation of the earliest subject studies. The authors are indebted to Nancy Thorson for assistance with subject recruitment, records and procedures and to Olga Matlega for assistance with manuscript preparation.

REFERENCES

Adler DD, Carson PL, Rubin JM, Quinn-Reid D. Doppler ultrasound color flow imaging in the diagnosis of breast cancer: Preliminary findings. *Ultrasound Med Biol* 1990; 16:553-559.

Bergonzi M, Calliada F, Corsi G, Passamonti C, Bontioli C, Motta F, Urani A. Role of echo-color Doppler in the diagnosis of breast diseases: Personal experience. *Radiol Med (Torino)* 1993; 85:120-123.

Blohmmer JU, Chaoui R, Schmalisch G, Bollmann R, Lau HU. Differential breast tumor diagnosis by comparing blood circulation of the tumor with the contralateral breast using color coded, pulsed Doppler ultrasound. *Geburtshilfe Frauenheilkd* 1995; 55:1-6.

Calliada F, Raieli G, Sala G, Conti MP, Bottinelli O, La Fianza A, Corsi G, Bergonzi M, Campani R. Doppler color-echo in the echographic evaluation of solid neoplasms of the breast: 5 years of experience. *Radiol Med (Torino)* 1994; 87:28-35.

Carson PL, Adler DD, Fowlkes JB, Harnist K, Rubin JM. Enhanced color flow imaging of breast cancer vasculature: Continuous wave doppler and 3-D display. *J Ultrasound Med* 1992; 11:377-385.

Carson PL, Li X, Pallister J, Moskalik A, Rubin JM, Fowlkes JB. Approximate quantification of detected fractional blood volume in the breast by 3D color flow and doppler signal amplitude imaging. In: Levy M, McAvoy BR, eds. 1993 Ultrasonics symposium proceedings. Piscataway, NJ: Institute for Electrical and Electronics Engineers. IEEE Cat. No.93CH3301-9. 1993:1023-1026.

Carson PL, Govil A, Moskalik AP, Roubidoux MA, Helvie MA, Adler DD, Fowlkes JB, Rubin JM. 3D Power Mode ROI Analysis for Breast Vascularity Assessment. Abstract booklet. In: Ninth International Congress on the Ultrasonic Examination of the Breast, Sept. 28, Indianapolis. 1995:30-31.

Chang CHJ, Nesbit DE, Fisher DR, Fritz SL, Dwyer SJ III, Templeton AW, Lin F, Jewell WR. Computer tomographic mammography using a conventional body scanner. *Am J Roentgenol* 1982; 138:553-558.

Chenevert TL, Helvie MA, Aisen AM, Francis IR, Adler DD, Roubidoux MA, Lunday F. Dynamic 3-D scanning by partial K-space sampling: Initial application for Gad-DTPA contrast enhancement rate characterization of breast lesions. *Radiology* 1994; 196:135-142.

Cosgrove DO, Kedar RP, Bamber JC, Al-Murrani B, Davey JBN, Fisher C, McKenna JA, Alsanjari NA. Breast disease: color Doppler US in differential diagnosis. *Radiology* 1993; 189:99-104.

Cosgrove DO, Lees WR. Abdominal ultrasound workshop. Eurodop: European Doppler Ultrasound Symposium. Oxford, UK: European Ultrasound Society, 1994.

De Albertis P, Oliveri M, Quadri P, Seratini G, Cavallo A, Orlando O, Perona F, Barile A, Gandolfo N. Retrospective analysis of color Doppler ultrasonography and flowmetry findings in solid nodular pathology of the breast. *Radiol Med (Torino)* 1995; 89:28-35.

Dixon JM, Walsh J, Paterson D, Chetty U. Color Doppler ultrasonography studies of benign and malignant breast lesions. *Br J Surg* 1992; 79:259-60.

Dymling SO, Persson HW, Hertz CH. Measurement of blood perfusion in tissue using Doppler ultrasound. *Ultrasound Med Biol* 1991; 15:433-444.

Fein M, Delorme S, Weisser G, Zuna I, van Kaick G. Quantification of color Doppler for the evaluation of tissue vascularization. *Ultrasound Med Biol* 1995; 21:1013-1019.

Folkman J, Shing J. Angiogenesis. *J Biol Chem* 1992; 267:11931-10934.

Goldberg BB, Taylor KJW, Merritt CM, Mendelson E, Madjar H, Bokobza J, Cosgrove D. Multicenter breast mass evaluation using imaging and color Doppler. Imaging: Applications and clinical results (abstract). *J Ultrasound Med* 1993; 60:S2:17.

Heilenkötter U, Jagella P. Color Doppler ultrasound of breast tumors needing excision: Presentation of an examination method. *Geburts Frauen* 1993; 53:247-252.

Huber S, DeJorme S, Knopp MV, Junkermann H, Zuna I, von Fourrier D, van Kaick G. Breast tumors: computer-assisted quantitative assessment with color Doppler US. *Radiology* 1994; 192:797-801.

Kedar R, Cosgrove D. Residual breast cyst mimicking a carcinoma on B-mode and color Doppler ultrasonography (letter). *J Ultrasound Med* 1994; 13:119-20.

Kedar RP, Cosgrove DO, Bamber JC, Bell DS. Automated quantification of color doppler signals: a preliminary study in breast tumors. *Radiology* 1995; 197:39-43.

Kuijpers TJ, Obdeijn AI, Kruyt RH, Oudkerk M. Solid breast neoplasms: differential diagnosis with pulsed Doppler ultrasound. *Ultrasound Med Biol* 1994; 20:517-520.

LaGalla R, Caruso G, Marasa L, D'Angelo I, Cardinale AE. Angiogenetic capacity of breast neoplasms and correlation with color Doppler semiology. *Radiol Med (Torino)* 1994; 85:392-5.

Leucht WJ, Rabe DR, Humbert K. Diagnostic value of different interpretative criteria in real-time sonography of the breast. *Ultrasound Med Biol* 1988; 14:59-73.

Madjar H, Prompeler HJ, Sauerbrei W, Wolfarth R, Pfleiderer A. Color Doppler flow criteria of breast lesions. *Ultrasound Med Biol* 1994; 20:849-858.

McNicholas MM, Mercer PM, Miller JC, McDermott EW, O'Higgins NJ, MacErlean DP. Color Doppler sonography in the evaluation of palpable breast masses. *Am J Roentgenol* 1993; 161:765-771.

Moskalik A, Carson PL, Meyer CR, Fowlkes JB, Rubin JB, Roubidoux MA. Registration of 3D compound ultrasound scans of the breast for refraction and motion correction. *Ultrasound Med Biol* 1995; 21:769-778.

Peters-Engl C, Medi M, Leodolter S. The use of colour-coded and spectral Doppler ultrasound in the differentiation of benign and malignant breast lesions. *Br J Cancer* 1995; 71:137-9.

Picot PA, Rickey DW, Mitchell R, Rankin RN, Fenster A. Three-dimensional Doppler imaging. *Ultrasound Med Biol* 1993; 19:95-104.

Pretorius DH, Nelson TR, Jaffe JS. 3-dimensional sonographic analysis based on color flow Doppler and gray scale image data: a preliminary report. *J Ultrasound Med* 1992; 11:225-232.

Rubin JM, Adler RA, Fowlkes JB, Spratt S, Pallister JE, Chen JF, Carson PL. Fractional moving blood volume estimation using Doppler power imaging. *Radiology* 1995; 197:183-190.

Rubin JM, Bude RO, Carson PL, Adler RS, Bree RL. Power Doppler: A potentially useful alternative to mean-frequency based color Doppler sonography. *Radiology* 1994; 190:853-856.

Sakkis S. Angiography of the female breast. *Ann Clin Res* 1974; 6 (Suppl. 12):1-47.

Scherzinger AL, Belgam RA, Carson PL, Meyer CR, Sutherland JV, Bookstein FL, Silver TM. Assessment of ultrasonic computed tomography in symptomatic breast patients by discriminant analysis. *Ultrasound Med Biol* 1989; 15:21-28.

Stavros AT, Thickman DL, Rapp CL, Dennis MA, Parker SH, Sisney GA. Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions. *Radiology* 1995; 196:123-134.

Swets JA, Pickett RM. Evaluation of diagnostic systems. New York: Academic Press, 1982:29-33.

Venta LA, Dudiak CM, Salamon CG, Flisak ME. Sonographic evaluation of the breast. *Radiographics* 1994; 14:29-50.

Wahl RL, Cody RL, Hutchins GD, Mudgett EE. Primary and metastatic breast carcinoma: Initial clinical evaluation with PET with

the radiolabeled glucose analogue 2-[F-18]-fluoro-2-deoxy-D-glucose. *Radiology* 1991;179: 765-70.

Watt AC, Ackerman LV, Windham JP, Shetty PC, Burke MW, Flynn MJ, Grodinsky C, Fine G, Wilderman SJ. Breast lesions: differential diagnosis using digital subtraction angiography. *Radiology* 1986;159:39-42.

Weidner N. Intratumor microvessel density as a prognostic factor in cancer. *Am J Pathol* 1995;147:9-19.

Weidner N, Folkman J, Pozza F, *et al*. Tumor angiogenesis: a new significant and independent prognostic indicator in early stage breast carcinoma. *J Nat Cancer Inst* 1992;84:1875-1887.

APPENDIX 2

**Cumulative Power Distribution Function:
Technique for Defining a Stable Intravascular Point for Normalizing
Fractional Moving Blood Volume Estimates with Power Doppler US**

**Jonathan M. Rubin, M.D., Ph.D., Ronald O. Bude, M.D., J. Brian Fowlkes, Ph.D.,
R. Steven Spratt, Ph.D*, Paul L. Carson, Ph.D., Ronald S. Adler, Ph.D, M.D.**

**University of Michigan Hospitals
Department of Radiology
1500 E. Medical Center Dr.
Ann Arbor, MI 48109**

*** Diasonics Corporation
2860 De La Cruz Blvd
Santa Clara, CA 95050**

**Funded in part by a grant from Diasonics Corporation, USPHS Grant RO1CA55076 from
National Cancer Institute, and US Army Medical Research Materiel Command under contract
NO. DAMD17-94-J-4144**

Address Correspondence to JMR

FX: 313-936-7948

e-mail: jrubin@umich.edu.

Abstract:

Purpose: The calculation of a fractional moving blood volume estimate (FMBVE) from a power Doppler signal is affected by variables such as tissue attenuation and depth, transducer aperture, and rouleaux formation. We present a method for normalizing the power Doppler signal to the expected signal from 100% blood.

Materials and Methods: In order to locate signal from flowing blood with a relatively consistent backscatter coefficient, we estimated the knee of the cumulative Doppler power distribution function (CDPDF). We used a flow tube phantom to test the ability of this knee to consistently locate a radial position at the periphery of the vessel, that would fall into a region of high shear stress and minimal rouleaux formation. We also studied how well the method normalized FMBVEs of the right renal cortex in a volunteer under conditions corresponding to different degrees of attenuation or different body habitus and in a group of six normal volunteers on three different occasions to estimate variability.

Results: Over a range of 5 flow velocities and receiver gains varying between undersaturated to severely oversaturated, the ratio of the calculated flow tube area compared to an observer defined gold standard was $89 \pm 7\%$. The best estimating technique for defining the knee of the CDPDF was within 0.3 dB of the subjectively selected knee. In humans, the technique effectively normalized the FMBVEs over an 8 dB receiver gain variation, and the mean and standard deviation of FMBVEs over the cohort of six volunteers was $37.6 \pm 3.6\%$.

Conclusion: Vascularity estimates using power Doppler ultrasound are feasible using a normalization scheme based on CDPDF. The method is robust and simple, and can be used to make vascularity estimates using power Doppler ultrasound.

Keywords: Ultrasound, Vascularity, Perfusion, Doppler, Power Doppler

Introduction:

Although of significant interest, ultrasound Doppler vascularity measurements have been employed on a variable basis in several organ systems (1-4). They have generally, but not exclusively, been based on counting color pixels in standard color Doppler scans and have been used clinically most frequently to assess tumor vascularity in the breast and prostate (5-13). The choice of these organs is informative, because it reveals one of the greatest shortcomings of ultrasound vascularity assessments, their depth dependence. In all of the examples in which quantitation has been employed, the target is either superficial or the transducer can be positioned very close to the area of interest. This is because vascularity is assessed by counting the number of color pixels in a given region, and anything that alters color pixel count will affect a vascularity assessment. Unfortunately, the density of blood vessels, the quantity one wishes to measure, is only one of the factors influencing color pixel count. Major contributors to this measurement are attenuation, aperture size, or any other property that alters the strength of backscattered echoes. Given identical tissue samples, their vascularity on ultrasound will change the deeper in the body the tissue is located. The closer a tissue is placed to the transducer, the more vascular it will appear. For example, flow in ovaries may be very hard to detect transabdominally, and yet very easily seen transvaginally. The vascularity detection improvement is due primarily to reduced attenuation path in the latter, yielding some combination of higher signal and allowing the use of higher frequency transducers (14). Because of this very definite attenuation affect, vascularity measurements using ultrasound are very tricky to make and interpret.

A potential solution to this problem does exist. If one can identify an object of known vascularity or blood volume, one could use the Doppler power in that object to normalize the values in the surrounding tissue (15-17). If the tissues of interest are near enough to the known object, the fraction of the Doppler power in the surrounding tissue relative to the standard can be used as a vascularity estimate compared to the standard. This

process would effectively compensate for any attenuation or transducer beam profile effects. Since blood vessels contain 100% blood, objects of known vascularity exist throughout the body, and could be used to normalize vascularity measurements. We will call such measurements fractional moving blood volume estimates (FMBVE) (17).

There is one problem, however, that has to be overcome before this method can be applied. The problem arises because the backscattered signal from blood is variable due mostly to the degree of rouleaux formation. (18-22). The reason for this is that rouleaux effectively increases the size of the scattering units in blood. Clumps of red blood cells will backscatter more ultrasound energy than will single individual cells, making areas of vessels with more rouleaux have stronger Doppler signals than those with less. This effect is most pronounced when comparing the Doppler power in the center of large blood vessels to the edges. There is typically much more signal amplitude in the center of vessels due to increased rouleaux formation (23). The higher shear rate and shear stress near the vessel wall applies more force to separate rouleaux clumps of red cells. The magnitude of the shear rate, which is directly proportional to the shear stress, depends on the velocity profile (24). For a parabolic flow profile, the shear rate, and hence shear stress, is directly proportional to the radius (20). Hence for a 1 cm diameter vessel, the shear rate at the wall will be 10 times that .5 mm from the center of the vessel. For near plug flow, the shear rate at the wall can be arbitrarily high. Because of this variation in the appearance of blood flow, one cannot just identify a blood vessel, obtain a normalization value, and calculate a FMBVE at a given depth. A more sophisticated technique is required.

It is known that smaller vessels have greater shear than do large vessels for a given mean velocity (25). Since we would like to normalize vascularity by the power corresponding to 100% blood in tissue, where the blood vessels are relatively small, a value outside the region of high rouleaux would be appropriate. Secondly because of a no slip boundary condition, there is no flow immediately adjacent to the wall of a vessel. Hence, the "correct"

normalizing value lies somewhere between the high value in the center of a vessel and the zero value at the wall.

In order to perform a FMBVE measurement, we will first calculate a cumulative power distribution function (CDPDF) over the region of interest (see below). We will then identify the knee of the CDPDF which will then be used in an algorithm to define a boundary point within highly vascular areas in which we will search for a normalization value. The knee of this distribution is a reasonable choice for segmenting a population into two groups similar to what is done in segmenting populations in receiver operator characteristic analysis (26,27). Once the normalization value is measured, we will use that value to determine the FMBVE in tissue. We will show that this technique can compensate for depth, attenuation, receiver gain, and machine variability. We will do this using flow phantoms and normal human volunteers. Ultimately, we believe that the results strongly suggest that this technique could be generally employed to assess vascularity in tissues.

Materials and Methods:

Definition:

The (CDPDF) is the integral of the histogram of the Doppler powers within an ROI (28). It is defined in a discrete sense as

$$N(p) = \sum_{i=0}^{i=p} n(i)$$

where $n(i)$ is the number of pixels with the i th power value and $N(p)$ is the total number of pixels with power values less than or equal to p , i.e. $i \leq p$. Because this is an integral of the histogram, the CDPDF has some very nice random noise suppressing properties that make it an interesting candidate for finding a normalization value. In a cumulative distribution, fluctuations are referenced to the entire display dynamic range, whereas with a standard

histogram, being the derivative of the cumulative distribution, fluctuations are seen relative to locally changing values for which any variation is a much larger percentage.

Phantom studies:

We designed a flow tube experiment to evaluate the boundary detection properties of the CDPDF. The tube consisted of an elongated balloon, 5.5 mm diameter, running through a fluid bath. The balloon was draped across a sound absorbing pad to suppress reverberations. Flow in the balloon was generated using a pulsatile pump (Harvard Apparatus, South Natick, MA) with a 70% duty cycle and 50 strokes per minute. Before reaching the balloon, the fluid passed through a long segment of compliant gum rubber tubing that made the flow non-pulsatile. Degassed water containing corn starch (0.5 g per liter of water) for scatterers was run through the balloon. In order to simulate real flow containing tissue, we put sephadex particles (20-80 μm diameter at concentrations of .0373 - .2372 g/liter depending on the background Doppler power level) into the bath outside the balloon. By stirring the bath, we could introduce a Doppler shift in the medium and generate a Doppler power in the background similar to the effect of blood-flow-containing tissue. The quantities of sephadex and corn starch were selected so that the tube always visually had more power than the water bath at each flow setting in order to simulate a vessel surrounded by flowing blood within perfused tissue. We made three measurements at each flow rate investigated (Table 1). These three measurements were made to simulate a weak signal deep in tissue corresponding to low gain, an optimal gain setting (see below) for tissue that was superficial enough or vascular enough to supply sufficient signal, and high gain in which the signal in the tube was saturated to simulate bad technique or the presence of contrast agents.

In order to define the gain settings for the flow tube measurements to test the CDPDF algorithm, we had to define what low gain, optimal gain, and high gain would be. We used Diasonics VST (Diasonics Corporation, San Clara, CA) ultrasound scanners to make the measurements. We used a 10 MHz linear array transducer with 6.0 MHz Doppler frequency.

The angle of the scanhead to the tube was 22° to generate a Doppler angle. We measured signal strength and saturation level by assigning a green tag value to Doppler power measurements. Each green tag step defined by the manufacturer corresponded to 0.4 dB. We set the saturation level such that the highest green tagged pixels in the flow tube were within 3 clicks steps of the highest dB level, 24 dB. This corresponds to the highest pixel values in the flow tube being no more than about 23 dB (23.2 dB) on the 24 dB scale, which represents dynamic range of the power mode signal. When the Doppler gain was set such that the pixels with the highest power in the image fell into this range, the settings were considered optimally saturated. We then set the saturation limit in a range that was clearly undersaturated (5-7 dB) below the optimally saturated setting, and finally set a range in which the Doppler power in the tube was clearly oversaturated (5-7 dB above optimal saturation). A representative image from each flow run was stored digitally to the hard disk on the scanner and transferred to an image analysis program (Advanced Visual Systems Inc., Waltham, MA) for analysis.

The boundary of 100% blood as seen by the ultrasound beam in the flow tube was identified using a cumulative distribution based on the entire data set and was compared to an observer traced perimeter of the tube which was considered the "gold standard" (Figure 1). To define the region over which the cumulative distribution was taken, an observer first traced the entire area of the stored image of the tube and the water bath excluding the sound absorbing rubber pad. The pad was excluded since having no "flow" it would bias the measurement by introducing a structure that was much different than the paradigm of interest, i.e. blood vessel and background tissue with flow. A cumulative distribution was calculated based on the power values within this enclosed region. The knee or point of maximum curvature of the distribution was then subjectively selected on the cumulative distribution curve for each flow measurement. We used the visual estimate of the knee of the distribution for this calculation since it was deemed to be the best estimator of that point. In order to compensate for variability, we selected a range of about ± 1 dB about the selected

threshold and calculated the areas within these limits. Of course, this method of detecting the threshold would not work in practice, and we, therefore, developed techniques for automating the location of the knee (see below). However for this part of the study, it was felt that our best estimate of the knee, the visual estimate, was the most appropriate to use.

An observer then traced the tube as a gold standard for comparison of the cumulative distribution selected boundary. The tube was traced three times at each flow setting, the area calculated with each traced perimeter, and a mean area calculated for the tube. The mean areas for the cumulative distribution detected region and the manually traced region were then compared.

Automated Distribution Determination:

In order to implement a fractional blood volume measurement *in vivo* and to make more objective measurements, we wanted to calculate in an automated way the position of the knee of the distribution. We employed two techniques, each being initiated in the same way. A best fit line was first plotted through the distribution using a standard least squares technique. For the first technique, the points of intersection of the line with the distribution were located and tangent lines were drawn parallel to the cumulative distribution curve at the points of intersection. The intersection of these two tangent lines defined the knee of the cumulative distribution. The points of intersection were determined by taking the difference between the cumulative distribution value and the calculated estimate using the best fit line for each successive power value. For those points at which the sign of the difference changed, there must be an intersection between the best fit line and the distribution between that point and its neighbor, either one of which could be defined as the intersection for the purposes defined here. The tangent lines were empirically defined over 21 point regions, the intersection point ± 10 points. (Figure 2) (Further details are in the appendix.)

The second method begins as the first, fitting the cumulative distribution with a best fit straight line. The distribution is then rotated about the best fit line so that the line falls onto the abscissa. Based on the shape of the cumulative distribution, this function should

have a negative curvature and there should therefore be a global maximum. A global maximum could then easily be identified and corresponded to the knee. In order to account for noise about the global maximum, a region of ± 10 points around this initially identified global maximum was constructed and fitted with a second order polynomial. The power value corresponding to the peak of this polynomial could also represent the knee of the distribution. Although not identical, this technique is somewhat similar to other methods that employ second order polynomials to search for maxima (29).

The tangent method and the curve fitting method, both the initial rotated peak or the second order polynomial fitted peak, were then compared to the subjective gold standard for finding knees of the cumulative distributions for the tube flow data by taking a mean absolute difference between the measurements.

Human Studies:

The human studies were divided into two phases. The first consisted of multiple scans of one volunteer at different gain settings to simulate different body habitus, while the second consisted of multiple studies over different normal volunteers to investigate variability.

In the first phase, the right kidney of a single healthy male volunteer (age 49 years, weight 76 kgms, height 173 cms) was scanned on five different occasions. The scans were performed using a 3.5 MHz curved linear scanhead using a PRF of 1000, a color Doppler frequency of 2.5 MHz, and an ensemble length of 10. The receiver gain was successively increased from 102 dB to 112 dB in 2 dB steps. Over this range, the images clearly progressed from highly undersaturated to highly oversaturated. This was manifested by lack of flow detection at the 102 dB setting to a very bright ballooned color display at the 112 dB setting. Again this range was used to simulate scanning a deep structure to the use of contrast agents. The typical optimal setting for power Doppler, i.e. when the noise floor just begins to break through the uniform color background (29), was at about 105-107 dB. During each sampling, the kidney was scanned longitudinally at each gain setting using a region of interest

(ROI) defined by the machine's color box that included as much of the near field cortex of the kidney as possible and at least half of the renal sinus. Because of the curvilinear shape of the color box implemented with this probe, some liver and perinephric fat were also included in the ROI. The image of the cortex chosen for analysis at any given scanning setting was the one that subjectively had the most visible blood flow. A cumulative distribution was taken over the entire color box and the knee of the distribution was automatically calculated with both the tangent and the curve fitting techniques described above. A green tag was assigned to all power values equal to or greater than the value identified as the knee of the distribution. The technique which identified a knee that assigned the greatest fraction of the color dynamic range to potentially 100% blood, i.e. green tagged the most pixels, was then used in the next stage of the algorithm.

A second cumulative distribution was then taken over only those points selected in the first iteration. The knee of this second distribution was then identified using the two techniques described above, and the value which assigned the largest fraction of the dynamic range to green was again used as the normalization value. This value became the normalization value for the FMBVE procedure or equivalently the lowest Doppler power corresponding to 100% blood value. This process is demonstrated in figure 3.

A region of cortex for which the FMBVE was to be estimated was then traced within the color box using the machine's tracker ball. Besides the selection of the size and location of the color box itself, this was the only portion of the examination in which there was operator interaction. A FMBVE was calculated from this region. The means and standard deviations of the selected initial green tag level were plotted as a function of receiver gain to show the effect of attenuation and potentially body habitus on the normalization level. Further, a mean value and standard deviation of FMBVE as function of receiver gain were plotted over the five measurement instances.

The second phase included studies of six healthy volunteers (one male, five females, ages 34 ± 6 years (range 27-41), weights 78 ± 22 kgms (range 57-116), heights 173 ± 11 cms

(range 160-188)} by one examiner (JMR). The volunteers were each scanned on three different days without preparation for purposes of practicality. The volunteers were scanned using the same machines employed above; optimizing the setting for power Doppler in the usual manner, i.e. the gain was increased sequentially until the noise floor just broke through the uniform background (30). Once the receiver gain was set, the same process for determining a FMBVE described for the single volunteer was used. The means and standard deviations for the FMBVE estimate for each of the volunteers were plotted.

Results:

Flow experiment.

The experimental flow runs are tabulated in Table 1. Comparing the areas of the tube enclosed using the knee of the cumulative distribution and the subjective boundary, the area of the tube detected using the cumulative distribution was $89 \pm 7\%$ of the subjective boundary. It must be said that at the highly saturated settings, the difference between the boundary and the background scatterers in the medium was hard to see, and the knee of the distribution was often felt to be a better marker of the boundary than the "gold standard". Further, it is obvious that there is blooming of the flow tube with increasing gain making the detected boundary move outward. This corresponded to a larger tube area for any given velocity (Table 1), and is to be expected with higher levels of saturation. This would manifest in vivo as enlargement of blood vessels with increasing gain.

We then compared the different knee detecting techniques to see how well they identified the knee compared to the subjectively chosen location (Table 2). It is clear that the tangent method underestimates the subjectively designated position of the knee, thereby assigning more power values to 100% blood. The mean absolute difference for the two tangent method from the knee was 0.91 dB. The second order polynomial fitting technique more closely approximated the subjectively selected knee, 0.33 dB for the peak value and 0.30 dB for the second order polynomial fitted peak.

Human studies:

The segmentation values of vascular structures from the relatively non-vascular background were plotted as a function of receiver gain (Figure 4). The plot shows an monotonic decline in the normalization level as a function of receiver gain. The slope is significantly different from zero, -0.845 ± 0.151 (95% confidence limits). This result is what would be expected if the knee of the cumulative distribution were compensating for overlying attenuation and depth of the organ, i.e. body habitus. The weaker the flow signal in general, the greater portion of the dynamic range that can be potentially assigned to 100% blood. Further, despite this change in normalization level with gain, the ultimate FMBVE measured is largely independent of receiver gain (Figure 5). This is based on taking a linear regression and determining the slope over the 8 dB highest dynamic range measurements, 104 - 112 dB, which is 0.00624 (-0.00067, 0.01315; 95% confidence limits). Thus zero is included in the 95% confidence limits. The addition of the lowest dynamic range measurement, 102 dB, produces a significant positive trend in the data, slope = 0.00798 (0.00198, 0.01400; 95% confidence limits). Finally, scans of the six volunteers shows the means and standard deviations of the FMBVEs one might expect with this technique (Figure 6). The standard deviation for any single individual can be up to ± 0.1 . This corresponds to approximately a 30% variation relative to the mean value. However, the mean of the means and standard deviation of the means for the six volunteers' FMBVEs is $37.6 \pm 3.6\%$.

Discussion:

The ultimate aim of many flow imaging techniques is to image vascularity or perfusion, and ultrasound is no exception. The opportunity for ultrasound to display vascularity has never been better. With contrast agents on the horizon and power Doppler, it is now possible for ultrasound imaging to directly display the amount of moving blood in tissue, since unlike standard mean frequency color Doppler, the power in the Doppler signal

is a reflection of the vascular blood volume of the tissue sample (20-22,31). Further, power Doppler has the potential to overcome a major limitation of ultrasound vascularity measurements. By normalizing power Doppler signals in tissue against a blood vessel that is in or near a region being evaluated, it is possible to compensate for distance from the transducer to the region of interest, attenuation coefficient of the overlying tissue, and the available imaging aperture among other things (32).

The problem with this strategy is that the amplitude of the Doppler shift signal is highly dependent on rouleaux formation (18-22), making it impossible to merely select any point within a large vessel for normalization. However, by selecting a normalization value from the region of high shear near the vessel wall, the affects of rouleaux can presumably be minimized. This would create a situation approximating that occurring in small vessels in tissue where the shear is high and rouleaux is relatively low (25).

Hence, a normalization technique must detect blood vessels and identify a position in the vessels to use for normalizing. The flow phantom study shows that the cumulative distribution can find blood vessels. In fact, it can optimally find them to within about 4% , .3 dB, compared to an observer. In the clinical portion of the study, we typically used the tangent estimate, the poorer of the two estimating techniques, because it tended to overestimate the number of pixels potentially containing 100% blood and thus increased the number of points included in the distribution. However, even using an optimal technique such as the subjective method in the flow phantom, regions are selected still in the "non-vascular" tissue bath that were identified as flow-containing (Figure 1). Similar findings occurred in the right kidneys studied. This is not surprising and would be expected, at least in the first distribution, given the presence of small blood vessels inside tissue which are analogous to the random high signals in the water bath.

We employed a continuous flow phantom even though flow *in vivo* can clearly be pulsatile. We decided to do this because power Doppler is generally highly frame averaged *in vivo*. This has the effect of averaging out the pulsatility in the sampling. Hence, the non-

pulsatile phantom would in some sense be an estimate of what happens in this first approximation. Yet, pulsatility will change the velocity distribution in a vessel which will change the power distribution based on shear rate effects as described above. This may require further study.

After determining the knee of the first distribution, we employed a second distribution to find the normalization value in vivo. The reason for this is that, as in the flow phantom, the first distribution essentially located the wall or the neighborhood just outside of the walls of blood vessels that would be used for normalization. We needed to find a value inside the vessels to which to normalize. Thus the second distribution over only the vessels accomplished this.

If we did not include a second distribution, we would have had the problem of distinguishing an organ of normal vascularity that is located deep in the body from a superficial organ with low vascularity. The reason for this is that the first distribution only basically locates the walls of large vessels as shown above. It will find the walls whether an organ is deep and normal or superficial and hypovascular. Since we use an approximation of the knee as the threshold and this knee is defined relative to the parenchymal, relatively non-vascular tissue, dividing this threshold into the tissue value will give a very similar ratio in both cases. The distinction between a deep organ and a superficial one does not lie in the vessel boundaries but is in Doppler power of the blood. For identical vessels, the more superficial, the more power the blood will have. This requires that a second distribution be taken over the vessels themselves to actually find a normalization value representative of the blood in the vessels. This value will be a function of the vessel depth. We could have just identified the vessels manually or by some other technique, and performed a distribution over the vessels only. Then we would have needed only one distribution. The first distribution lets us automate the method.

There obviously have to be at least two points of negative curvature in order to find two knees in our process, and it makes sense that there should be at least two. The first knee

separates the relatively avascular parenchyma from the blood vessels, and the second knee separates the high rouleaux central portions of blood vessels from the high shear margins of the vessels. There will always be more points in the first distribution, since it includes the vessels and the relatively avascular tissue, and given a large enough region of interest, the first knee will dominate and will be the one selected by our method. Once the vessels have been segmented out, the method will detect the second knee. If number of points included in this second distribution is small, it may limit the ability to accurately detect a knee in some circumstances. We did not encounter this problem in this study; however, this may require a change in strategy when very small regions are being measured. This effect can be seen in Figure 3b and c where the knee is very well defined in 3b, but not as well define, although present, in 3c.

It is clear in this study, that we have not determined where that precise site in the vessel is or what the true normalization value should be. However, we have done the next best thing, i.e. identified a method that compensates for the rouleaux and attenuation problem in a general way. First, we have shown this since the plot of the initial normalization value with receiver gain showed a continuous compensation for different amounts of signal by setting the normalization value to different fractions of the usable color dynamic range (Figure 4). This change in normalization compensated for the varying amounts of backscattered signal received at the different receiver gains which corresponds to increasing amounts of attenuation, increasing depth, and/or decreasing amounts of available transducer aperture. The slope of the regression line through the data defined in Figure 4 shows a clear trend with a slope whose 95% confidence limits did not intersect zero. Thus there was a distinct and significant decrease in the proportion of the color dynamic range that is assigned to 100% blood as the signal strengthened, or conversely the weaker the signal, the more of the dynamic range that is assigned to 100% blood.

This result makes sense if one considers that if there were no detectable flow, flowing blood would look like everything else and the entire dynamic range would look like it

contained 100% blood. The same problem would also occur at highly oversaturated images where the gain is so high that everything, including noise, is enhanced to look like blood. This would create a jump at the high receiver gain end of the curve if we had gone high enough. However, at that point, everything would be saturated and it would be impossible to distinguish blood vessels from less vascular surrounding tissue. This is unlikely to happen in practice because the user would immediately recognize it and turn the gain down as occurs with blooming with contrast agents (33,34). However, the differential blooming effect of power Doppler is an advantage here where, unlike with mean frequency color, the display increase corresponds to the amount of signal present. Hence vessels will still look brighter than background tissue even with blooming (33,34), and our results suggest that as long as tissue can be distinguished from large blood vessels, the FMBVE method could work in principle.

Secondly, once depth dependence of the transducer beam profile and attenuation are compensated for, the measurement of the same tissue region should be identical if the normalization method is working. Figure 5 shows this to largely be the case. The slope of the regression line through the data describing the FMBVEs for gain settings 104-112 dB was not different from zero to within 95% confidence (35). Hence when there was sufficient signal to detect, the method compensated for the attenuation and transducer effects as one would hope. However, when the data included the 102 dB values, there was a significant trend. In addition, the variability in the 102 dB measurement was higher than any of the others. The implication here is that when there is very little flow to detect, the method under compensates and becomes more variable. This suggests that when the signal is weak and is harder to detect, the method tends to locate a normalization value within the high rouleaux portion of the flow stream and would, in a sense, overcompensate for the lost signal in the cortex where the rouleaux effects are small. This would give a low estimate because the method would be comparing high rouleaux areas in vessels to the low rouleaux areas in the tissue. At some attenuation level, the Doppler signal will be too weak to measure and any

normalization technique will fail. One compensating fact is that when the signal gets very weak, it is still possible to compare the measurement to itself (17). Hence, the normalization from day to day may make it possible to evaluate relative flow changes in the same organ even if the attenuation is high.

Finally, the evaluation of the volunteers without preparation estimated the potential variability in the method under scanning conditions. Clearly, one would like a technique with little or no variability at all, yet in some cases the standard deviation in the measurements could be up to 10%, producing a variability relative to the local mean of approximately 30% from day to day even when the signal was optimized. In these studies, there was no control for hydration, time of day, or fasting; all of which could have influenced the variability. Further, given the single volunteer studies where the variance is smaller, it may be reasonable to presume that the variability is the true variation in the blood flow to the renal cortex. It is clear to anyone who does sonography that there are regions of the cortex with more flow than others, so when scanning the same person on different days it is possible to obtain different results.

Will such variability preclude using this test? We do not think so. First, the mean values are very close among all the individuals suggesting that a mean value may be usable in any given case. Thus taking the mean of several measurements may give very reasonable results. This is exactly what is done presently with all RI measurements. Secondly, there can be equally wide variability among other parametric tests such as resistive indices, pulsatility indices, acceleration, and acceleration times (36-38). Each of these has found use in the medical community. Finally, there are ways to improve the collection on the horizon, i.e. 3D (16). This method would be perfectly integrated into a 3D acquisition system. The segmentation would be employed the same way, but the sampling would be more robust. Given the large set of flows that would be sampled in the cortex in 3D, the flow variation in an organ would likely decrease.

Besides the limitations described above, there are others. As mentioned above, we have not defined the absolute fractional blood volume in the kidney. Although the mean values obtained are within or close to the published normal levels in the kidney (39-41), we do not know if those numbers measured here are in fact absolute fractional blood volume estimates. This may be true, but it will likely have little effect on clinical application of such a measure as this. There are plenty of "indices" that relate to physiological parameters of interest, but may not directly measure them such as RI (42,43). Secondly, there are clearly flows that will not be included in this measurement. Capillary flow, for instance, is too slow to measure by present Doppler techniques. Thus, these vessels will not be included in any vascularity estimate using our method. Of course, capillaries will not contribute to any ultrasound color Doppler vascularity measure, mean frequency color flow or power mode. In addition, vascularity assessments using color Doppler techniques can not be easily normalized, unlike ours. Hence, we believe that our method should surpass these.

In conclusion, we have defined a method for normalizing power Doppler FMBVE using properties of the cumulative Doppler power distribution. The technique is almost completely non-subjective and could easily be automated. The method works in phantom studies to find blood vessel boundaries, can be normalized for depth, gives constant measurements for the same flow situation largely independent of body habitus, and gives reasonable estimates of mean FMBVE in normal volunteers with a variability similar to other functional parametric indices. Further studies in pathological cases are needed to define its utility.

REFERENCES

1. Tanaka S, Kitamura T, Fujita M, et al. Color Doppler flow imaging of liver tumors. *AJR* 1990;154:509-514.
2. Shiamato K, Sadayuke S, Ishigaki T, et al. Intratumoral blood flow: evaluation with color Doppler echography. *Radiology* 1987;165:683-685.
3. Bourne T, Campbell S, Steer C, et al. Transvaginal colour flow imaging: a possible new screening technique for ovarian cancer. *Br Med J* 1989;299:1367-1370.
4. Luker GD, Siegel MJ. Pediatric testicular tumors: evaluation with gray-scale and color Doppler US. *Radiology* 1994;191:561-564.
5. Adler DD, Carson PL, Rubin JM, Quinn-Reid D. Doppler ultrasound color flow imaging in the study of breast cancer: preliminary findings. *Ultrasound Med Biol* 1990; 16:553-559.
6. Sohn CH, Stolz W, Grischke EM, et al. Die dopplersonographische untersuchung von mammatumoren mit hilfe der farbdopplersonographie, der duplex sonographie und des CW-Dopplers. *Zentralbl Gynakol* 1992;114:249-253.
7. Dixon JM, Walsh J, Paterson D, Chetty U. Colour Doppler ultrasonography studies of benign and malignant breast lesions. *Br J Surg* 1992;79:259-260.
8. Cosgrove DO, Bamber JC, Davey JB, McKinna JA, Sinnett HD. Color Doppler signals from breast tumors. work in progress. *Radiology* 1990;176:175-180.

9. Cosgrove DO, Kedar RP, Bamber JC, et al. Breast diseases: color Doppler US in differential diagnosis. *Radiology* 1993;189:99-104.
10. Kedar RP, Cosgrove DO, Bamber JC, Bell DS. Automated quantification of color Doppler signals: a preliminary study in breast tumors. *Radiology* 1995;197:39-43.
11. Rifkin MD, Sudakoff GS, Alexander AA. Prostate: techniques, results, and potential applications of color Doppler US scanning. *Radiology* 1993;186:509-513.
12. Newman JS, Bree RL, Rubin JM. Prostate cancer: diagnosis with color Doppler sonography with histologic correlation of each biopsy site. *Radiology* 1995;195:86-90.
13. Carson PL, Moskalik AP, Govil A, et al. The 3D and 2D color flow display of breast masses. *Ultra Med Bio* 1997 (in press).
14. Shung KK, Sigelmann RA, Reid JM. Scattering of ultrasound by blood. *IEEE Trans Biomed Eng* 1976;BME-23:460-467.
15. Hottenger CF, Meindl JD. Blood flow measurement using the attenuation-compensated flowmeter. *Ultrasonic Imaging* 1979;1:1-15.
16. Carson PL, Li X, Pallister J, Moskalik A, Rubin JM, Fowlkes JB. Approximate quantification of detected fractional blood volume in the breast by 3D color flow and Doppler signal amplitude imaging, in Levy M, McAvoy BR, eds. 1993 ultrasonics symposium proceedings. Piscataway, NJ:Institute for Electrical and Electronics Engineers, IEEE Cat #93CH3301-9, pp 1023-1026.

17. Rubin JM, Adler RS, Fowlkes JB, et al. Fractional moving blood volume: estimation with power Doppler US. *Radiology* 1995;197:183-190.
18. Sigel B, Machi JJ, Beitler JC, Justin JR. Red cell aggregation as a cause of blood-flow echogenicity. *Radiology* 1983;148:799-802.
19. Machi JJ, Sigel B, Beitler JC, Coelho JCU, Justin JR. Relation of in vivo blood flow to ultrasound echogenicity. *JCU* 1983;11:3-10.
20. Shung KK, Cloutier G, Lim CC. The effects of hematocrit, shear rate, and turbulence on ultrasound Doppler spectrum from blood. *IEEE Trans Biomed Eng* 1992;39:462-469.
21. Shung KK. In vitro experiment results on ultrasonic scattering in biological tissues. In: Shung KK, Thieme GA, eds. *Ultrasonic scattering in biological tissues*. Boca Raton, Fl: CRC Press, 1993;291-312.
22. Shung KK, Kuo IY, Cloutier G. Ultrasound scattering properties of blood. In: Roelandt J, Gussenhoven EJ, Bom N, eds. *Intravascular ultrasound*. Dordrecht, the Netherlands: Kluwer Academic Publishers, 1993;119-139.
23. Yuan YW, Shung KK. Ultrasonic backscatter from flowing whole blood. I: Dependence on shear rate and hematocrit. *J. Acoust Soc Am* 1988;84:52-58.
24. Feynman RP, Leighton RB, Sands M. *The Feynman lectures on physics*. Vol II. Reading, MA: Addison-Wesley, 1964;41-2.

25. Whitmore RL. The flow fluids. In: Whitmore RL. ed. *Rheology of the circulation*. Oxford: Pergamon Press, 1968;37-46.
26. Metz CE. Basic principles of ROC analysis. *Sem Nuc Med* 1978;8:283-298.
27. Halperin EJ, Albert M, Krieger AM, Metz CE, Maidment AD. Comparison of receiver operating characteristic curves on the basis of optimal operating points. *Acad Radiol* 1996;3:245-253.
28. Papoulis A. *Probability, random variables, and stochastic processes*. New York: McGraw-Hill, 1965;92-94.
29. Press WH, Flannery BP, Teukolsky SA, Vetterling WT. *Numerical recipes in C. The art of scientific computing*. Cambridge, England: Cambridge University Press, 1988;299-302.
30. Rubin JM, Bude RO, Carson PL, Bree RL, Adler RS. Power Doppler US: a potentially useful alternative to mean frequency-based color Doppler US. *Radiology* 1994;190:853-856.
31. Meyerowitz CB, Fleischer AC, Pickens DR, et al. Quantification of tumor vascularity and flow with amplitude color Doppler sonography in an experimental model: preliminary results. *J Ultra Med* 1996;15:827-834.

32. Goldstein A. Physics of Ultrasound. In: Rumack CM, Wilson SR, Charboneau JW, eds. *Diagnostic ultrasound*. St. Louis: Mosby-Year Book, 1991; 2-18.
33. Murphy KJ, Bude RO, Dickinson LD, Rubin JM. Use of intravenous contrast in transcranial sonography: preliminary observations. *Academic Rad* 1997;4:577-582.
34. Burns PN. Harmonic imaging with ultrasound contrast agents. *Clinical Radiology* 1996; 51 suppl. 1:50-55.
35. Metz CE. Quantification of failure to demonstrate statistical significance. The usefulness of confidence intervals. *Invest Radiol* 1993;28:59-63.
36. Kliewer MA, Tupler RH, Carroll BA, et al. Renal artery stenosis: analysis of Doppler waveform parameters and tardus-parvus pattern. *Radiology* 1993;189:779-787.
37. Keogan MT, Kliewer MA, Hertzberg BS, DeLong DM, Tupler RH, Carroll BA: Renal resistive indexes: variability in Doppler US measurement in a healthy population. *Radiology* 1996;199:165-169.
38. Paulson EK, Kliewer MA, Frederick MG, Keogan MT, Delong DM, Nelson RC. Hepatic artery: variability in measurement of resistive index and systolic acceleration time in healthy volunteers. *Radiology* 1996;200:725-729.
39. Selkurt EE. The renal circulation. In: Hamilton WF, Dow P, eds. *Handbook of physiology, circulation*. Vol 2, Sect 2. Washington , DC: American Physiological Society, 1963:1471.

40. Ladefoged J, Pedersen F. Renal blood flow, circulation times and vascular volume in normal man measured by the intraarterial injection-external counting technique. *Acta Physiol Scand* 1967;69:220-229.
41. Wiencek JG, Feinstein SB, Walker R, Aronson S. 1993. Pitfalls in quantitative contrast echocardiography: the steps to quantitation of perfusion. *J Am Soc Echocardiogr* 6: 395-416.
42. Bude RO, Rubin JM, Platt JF, Fechner KP, Adler RS. Pulsus tardus: its cause and potential limitations in detection of arterial stenosis. *Radiology* 1994;190:779-784.
43. Halpern EJ, Deane CR, Needleman L, Merton DA, East SA. Normal renal artery spectral Doppler waveform: a closer look. *Radiology* 1995;196:667-673.

APPENDIX:

Quantitatively, the definition of the intersection points and tangent lines look like this:

$P'(x) = Ax + B$ is the form of the best fit line for a given distribution where $P'(x)$ is the estimated cumulative value for a given x , the Doppler power. If $P(x)$ is the true distribution from the data, then $P(x) - P'(x)$ will be zero at points of intersection of the true distribution and the best fit line. Since this is real data, we cannot rely on the actual points of intersection being in the data set. Hence, one can identify the closest point in the data set to the intersection as being the one where there is a sign change. For example an intersection would lie immediately between points x_i and x_{i+1} if $(P(x_i) - P'(x_i)) < 0$ and $(P(x_{i+1}) - P'(x_{i+1})) > 0$. As a first approximation one could chose either x_i or x_{i+1} as the intersection. Since we are only interested in a set of points centered at either x_i or x_{i+1} defining a tangent, it doesn't really matter which point we pick. Choosing either one of these points at each intersection, we empirically selected a 21 point region of interest centered about each selected point, e.g. $x_i \pm 10$ points, and fitted a line through each group of 21 points. We selected a 21 point region based on the fact that worked well empirically. This local fitted line is the local tangent.

The final complication is that it is possible that there could be more than two intersections of the global best fit line with the distribution. To get around this possibility, we progressively identified each point of intersection of the best fit line with the cumulative distribution moving from the origin toward increasing abscissa values, i.e. Doppler power. We then selected the first pair of lines in which the difference between their slopes was negative, i.e. their slopes were decreasing or the curvature was negative. For example if T is a tangent line and $i = \{\text{set of abscissa values of intersection points ranging sequentially from 1 to } n\}$, then $T(i)$ is the tangent line to the cumulative distribution based around i . If Slope $T(i)$ is the slope of tangent line $T(i)$ and if $T(i)$ and $T(i+1)$ are the first pair of tangent lines encountered moving towards the right from the origin for whose slopes $(\text{Slope } T(i) - \text{Slope } T(i+1)) < 0$, then the intersection of $T(i)$ and $T(i+1)$ is the chosen point for the knee.

Figure captions:

Figure 1: Cross-sectional image of corn starch containing flow tube (T) in water bath (B) containing sephadex particles. The water bath was manually stirred to generate a Doppler signal, represented in yellow, that simulated viable, blood perfused tissue. The tube, which represents a blood vessel, corresponds to 100% blood and, hence, has more Doppler power than the water bath corresponding to less vascular tissue. This is demonstrated by the fact that the tube is brighter in power mode than the water bath. The outer margin of the blue line outlining the flow tube marks the boundary of the tube as calculated using the automated boundary detector. The blue line is actually thicker than the true boundary threshold for purposes of display. The blue pixels equal the set of all pixels in the image such that $\text{threshold} \leq \text{pixel value} \leq \text{threshold} + 0.9 \text{ dB}$. Notice that there are regions in the water bath that are also either surrounded by blue or are blue colored. These would correspond to potential areas of 100% blood in tissue. The black area behind the tube and water bath is the top of the sound absorbing rubber pad. Also note that there is a blue ring in the color wheel in the upper lefthand corner (arrow). The area of the wheel that is located between the blue ring and the outer edge represents the portion of the entire power dynamic range that is assigned to 100% blood in this particular trial.

Figure 2: Plot of cumulative distribution for one of the phantom flow experiments. The plot shows the three best fit lines used in the two tangent technique. The global best fit line (arrow with solid head) intersects the distribution at two points. A tangent line is defined at each intersection point, and their intersection (arrow with hollow head) defines the position of the knee. This technique clearly underestimates the position of the knee. This particular plot was for one of the undersaturated experiments, which is why it does not extend to 24 dB.

Figure 3: a) Longitudinal image of the renal cortex with near-field hilar vessels in a volunteer. The brighter the orange, the higher is the Doppler power. Blue represents the noise floor or the lowest Doppler power displayed. Notice the large blue area in the near-field corresponding to liver tissue (L). b) Same longitudinal image as in (a) after a cumulative distribution has been taken over the entire color box in (a). The green tagged pixels correspond to those that could represent 100% blood based on the knee of the cumulative distribution plot shown in the lower left hand corner of the image. The cumulative distribution for the region of interest defined by the color box is shown in the lower left-hand corner. This curve is typical of the initial in vivo distributions performed in this study. The calculated position of the knee is marked with an arrow. All pixels with power values equal to or greater than the knee are colored green in the image. (The highlighted boundaries adjacent to the green are artifacts of photography and lie outside of the green tagged region.) c) This image is identical to 3a and 3b but shows a second group of pixels as labeled green. These pixels are a subset of those shown in figure 3b and represent the final pixels determined to contain 100% blood. They were selected based on the cumulative distribution plot, shown in the lower left-hand corner of this image, which was taken over only those pixels labeled green in figure 3b. The calculated position of the knee (arrowhead) corresponds to the lowest Doppler power representing 100% blood, and hence this value is used to normalize all the tissue values in the image. Note that the second distribution demonstrates a much truer appearing vascular geometry since it has segmented high shear rate from low shear rate blood flow within large vessels. Small vessels in tissue will not show up due to high shear and partial volume with tissue, i.e. a FMBVE of less than 100%. (The highlighted pixels adjacent to the green tagged regions are artifacts of photography and are not included in the green tagged region.) d) Same longitudinal image with the region of the cortex to be measured enclosed by the dotted line. The FMBVE was calculated over that entire region and is normalized by the value selected in 3c. The FMBVE in this case was

37.1% (arrowhead). (Again, the highlighted areas are artifacts and lie outside of the green tagged region.)

Figure 4: Plot of the green tag threshold value for the first cumulative distribution as a function of the scanner's receiver gain. These values represent the threshold for those pixels that may contain 100% blood. Only those pixels with higher power values than the threshold will be considered as possibly containing 100% blood in the algorithm employed. The points are bounded above and below by \pm one standard deviation. Notice that the slope is negative meaning that the lower the gain, which corresponds to weak signal or large amounts of attenuation, the more dynamic range is potentially assigned to blood.

Figure 5: Plot of the FMBVEs as a function of receiver gain for the single volunteer. The points represent means \pm one standard deviation.

Figure 6: Plot of FMBVEs for the six volunteers. The points represent means \pm one standard deviation.

TABLE CAPTIONS:

Table 1: The raw data from the 15 flow experiments. Gain corresponds to the receiver gain setting used and fractional area is the area of the tube calculated using the knee of the cumulative distribution over the image divided by the area subjectively determined by tracing the outline of the tube.

Table 2: Comparison of the different methods for locating the knee position of the cumulative distribution for the flow tube experiments. The subjective knee position is the gold standard. The other methods, the two tangents, the rotated curve peak, and the second order polynomial fitted peak, are described in the text.

Figure 1: Cross-sectional image of corn starch containing flow tube (T) in water bath (B) containing sephadex particles. The water bath was manually stirred to generate a Doppler signal, represented in yellow, that simulated viable, blood perfused tissue. The tube, which represents a blood vessel, corresponds to 100% blood and, hence, has more Doppler power than the water bath corresponding to less vascular tissue. This is demonstrated by the fact that the tube is brighter in power mode than the water bath. The outer margin of the blue line outlining the flow tube marks the boundary of the tube as calculated using the automated boundary detector. The blue line is actually thicker than the true boundary threshold for purposes of display. The blue pixels equal the set of all pixels in the image such that $\text{threshold} \leq \text{pixel value} \leq \text{threshold} + 0.9 \text{ dB}$. Notice that there are regions in the water bath that are also either surrounded by blue or are blue colored. These would correspond to potential areas of 100% blood in tissue. The black area behind the tube and water bath is the top of the sound absorbing rubber pad. Also note that there is a blue ring in the color wheel in the upper left-hand corner (arrow). The area of the wheel that is located between the blue ring and the outer edge represents the portion of the entire power dynamic range that is assigned to 100% blood in this particular trial.

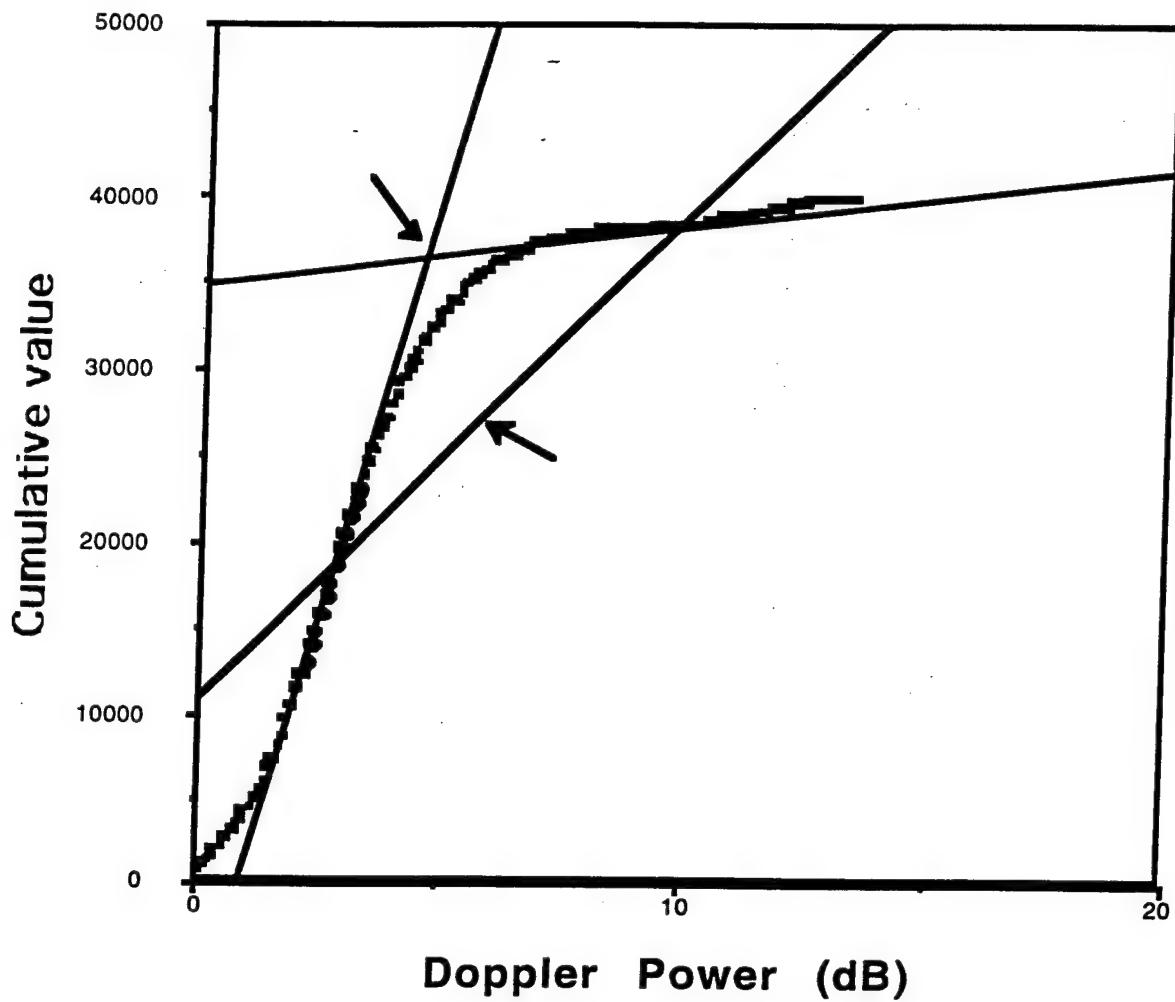


Figure 2: Plot of cumulative distribution for one of the flow phantom experiments. The plot shows the three best fit lines used in the two tangent technique. The global best fit line (arrow with solid head) intersects the distribution at two points. A tangent line is defined at each intersection point, and their intersection (arrow with hollow head) defines the position of the knee. This technique clearly underestimates the position of the knee. This particular plot was for one of the undersaturated experiments, which is why it does not extend to 24 dB.

Figure 3: a) Longitudinal image of the renal cortex with near-field hilar vessels in a volunteer. The brighter the orange, the higher is the Doppler power. Blue represents the noise floor or the lowest Doppler power displayed. Notice the large blue area in the near-field corresponding to liver tissue (L).

Figure 3: b) Same longitudinal image as in (a) after a cumulative distribution has been taken over the entire color box in (a). The green tagged pixels correspond to those that could represent 100% blood based on the knee of the cumulative distribution plot shown in the lower left hand corner of the image. The cumulative distribution for the region of interest defined by the color box is shown in the lower left-hand corner. This curve is typical of the initial *in vivo* distributions performed in this study. The calculated position of the knee is marked with an arrow. All pixels with power values equal to or greater than the knee are colored green in the image. (The highlighted regions adjacent to the green are artifacts of photography and lie outside of the green tagged region.)

Figure 3: c) This image is identical to 3a and 3b but shows a second group of pixels as labeled green. These pixels are a subset of those shown in figure 3b and represent the final pixels determined to contain 100% blood. They were selected based on the cumulative distribution plot, shown in the lower left-hand corner of this image, which was taken over only those pixels labeled green in figure 3b. The calculated position of the knee (arrowhead) corresponds to the lowest Doppler power representing 100% blood, and hence this value is used to normalize all the tissue values in the image. Note that the second distribution demonstrates a much truer appearing vascular geometry since it has segmented high shear rate from low shear rate blood flow within large vessels. Small vessels in tissue will not show up due to high shear and partial volume with tissue, i.e. a FMBVE of less than 100%. (The highlighted pixels adjacent to the green tagged regions are artifacts of photography and are not included in the green tagged region.)

Figure 3: d) Same longitudinal image with the region of the cortex to be measured enclosed by the dotted line. The FMBVE was calculated over that entire region and is normalized by the value selected in 3c. The FMBVE in this case was 37.1% (arrowhead). (Again, the highlighted areas are artifacts and lie outside of the green tagged region.)

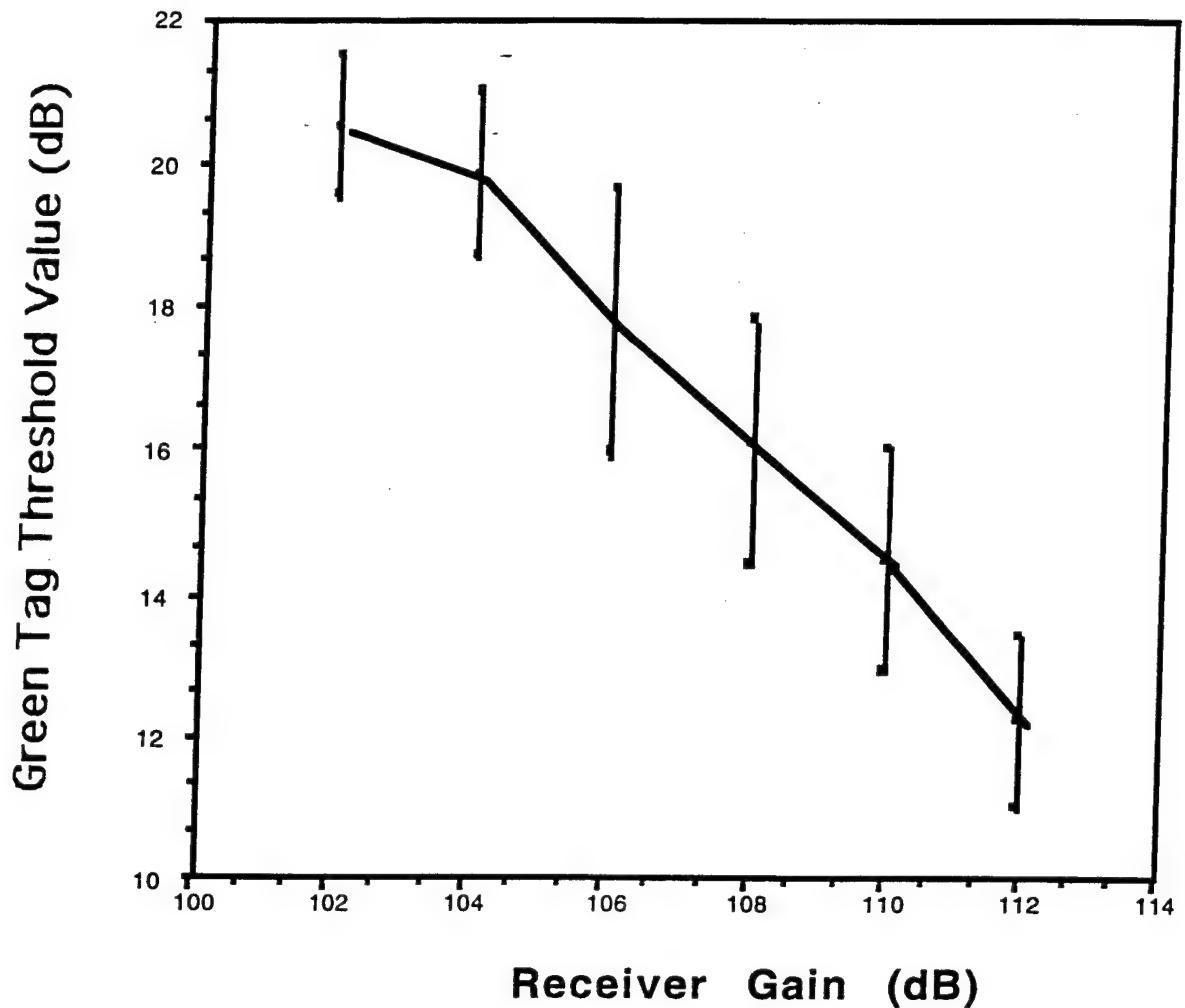


Figure 4: Plot of the green tag threshold value for the first cumulative distribution as a function of the scanner's receiver gain. These values represent the threshold for those pixels that may contain 100% blood. Only those pixels with higher power values than the threshold will be considered as possibly containing 100% blood in the algorithm employed. The points are bounded above and below by \pm one standard deviation. Notice that the slope is negative meaning that the lower the gain, which corresponds to weak signal or large amounts of attenuation, the more dynamic range is potentially assigned to blood.

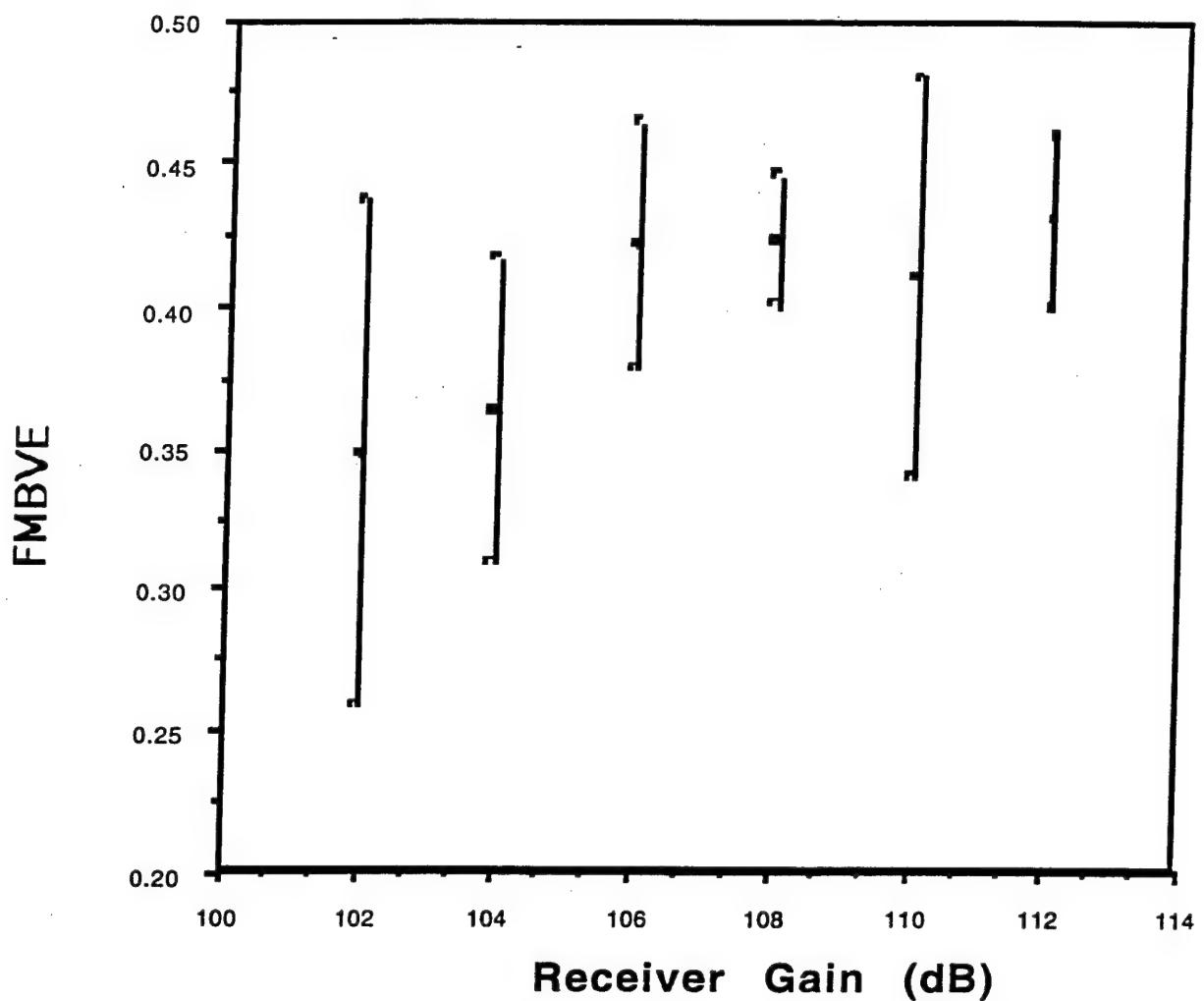


Figure 5: Plot of the FMBVEs as a function of receiver gain for the single volunteer. The points represent means \pm one standard deviation.

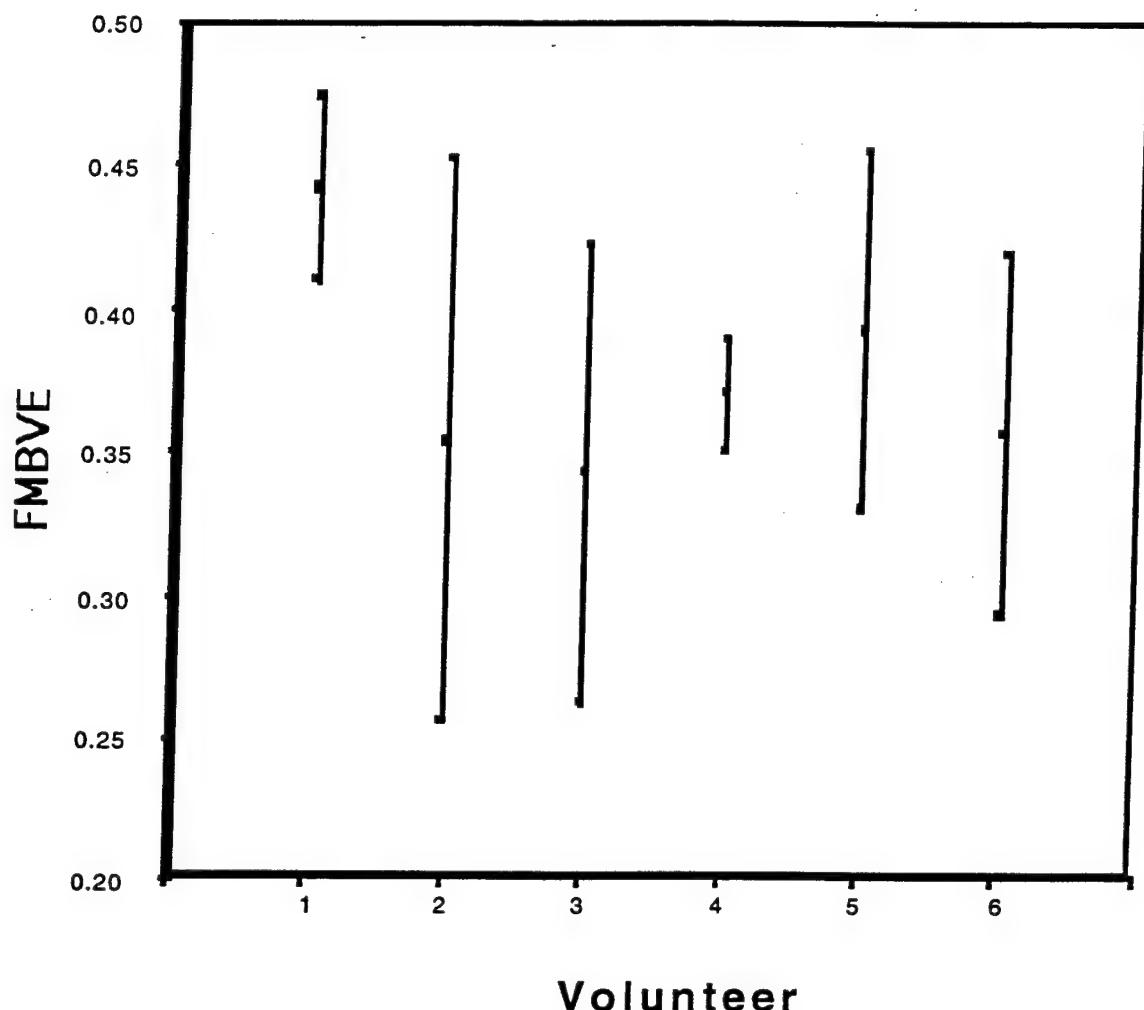


Figure 6: Plot of FMBVEs for the six volunteers. The points represent means \pm one standard deviation.

TABLE 1

Run	Mean Flow Velocity (cm/sec)	PRF (Hz)	Gain (dB)	Fractional Area
1	17	1250	79	0.849
2	17	1250	86	0.972
3	17	1250	92	0.875
4	35	2250	80	0.941
5	35	2250	85	0.893
6	35	2250	90	1.03
7	51	2750	84	0.894
8	51	2750	89	0.889
9	51	2750	94	0.93
10	60	3500	84	0.704
11	60	3500	89	0.917
12	60	3500	94	0.885
13	79	4500	84	0.903
14	79	4500	89	0.897
15	79	4500	94	0.886

Table 1: The raw data from the 15 flow experiments. Gain corresponds to the receiver gain setting used and fractional area is the area of the tube calculated using the knee of the cumulative distribution over the image divided by the area subjectively determined by tracing the outline of the tube.

TABLE 2

RUN	SUBJECTIVE KNEE POINT (dB)	TWO TANGENT MODEL (dB)	ROTATED CURVE PEAK (dB)	SECOND ORDER POLYNOMIAL FITTED PEAK (dB)
1	6.09	4.69	5.53	5.44
2	9.09	8.06	8.72	8.53
3	10.31	9.19	9.94	9.94
4	5.16	4.41	5.44	5.25
5	7.50	6.38	7.50	7.13
6	9.09	7.50	9.19	8.91
7	6.47	5.81	6.19	6.28
8	8.72	9.28	9.47	9.38
9	9.00	7.78	8.91	8.72
10	8.72	8.34	8.25	8.16
11	7.78	7.03	8.34	7.88
12	9.66	9.28	9.47	9.47
13	5.72	5.16	5.81	5.72
14	6.75	5.91	6.94	6.75
15	9.84	8.63	9.19	9.56

Table 2: Comparison of the different methods for locating the knee position of the cumulative distribution for the flow tube experiments. The subjective knee position is the gold standard. The other methods, the two tangents, the rotated curve peak, and the second order polynomial fitted peak, are described in the text.

Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations

Charles R. Meyer^{1*}, Jennifer L. Boes¹, Boklye Kim¹, Peyton H. Bland¹, Kenneth R. Zasadny², Paul V. Kison², Kenneth Koral², Kirk A. Frey² and Richard L. Wahl^{1,2}

¹Department of Radiology, 3307 Kresge III Research Building, University of Michigan Medical School, Ann Arbor, MI 48109, USA

²Department of Internal Medicine, Division of Nuclear Medicine, University of Michigan Medical School, Ann Arbor, MI 48109, USA

Abstract

This paper applies and evaluates an automatic mutual information-based registration algorithm across a broad spectrum of multimodal volume data sets. The algorithm requires little or no pre-processing, minimal user input and easily implements either affine, i.e. linear or thin-plate spline (TPS) warped registrations. We have evaluated the algorithm in phantom studies as well as in selected cases where few other algorithms could perform as well, if at all, to demonstrate the value of this new method. Pairs of multimodal gray-scale volume data sets were registered by iteratively changing registration parameters to maximize mutual information. Quantitative registration errors were assessed in registrations of a thorax phantom using PET/CT and in the National Library of Medicine's Visible Male using MRI T2-/T1-weighted acquisitions. Registrations of diverse clinical data sets were demonstrated including rotate-translate mapping of PET/MRI brain scans with significant missing data, full affine mapping of thoracic PET/CT and rotate-translate mapping of abdominal SPECT/CT. A five-point thin-plate spline (TPS) warped registration of thoracic PET/CT is also demonstrated. The registration algorithm converged in times ranging between 3.5 and 31 min for affine clinical registrations and 57 min for TPS warping. Mean error vector lengths for rotate-translate registrations were measured to be subvoxel in phantoms. More importantly the rotate-translate algorithm performs well even with missing data. The demonstrated clinical fusions are qualitatively excellent at all levels. We conclude that such automatic, rapid, robust algorithms significantly increase the likelihood that multimodality registrations will be routinely used to aid clinical diagnoses and post-therapeutic assessment in the near future.

Keywords: automatic, fusion, joint entropy, maximization, minimization, multimodality, mutual information, optimization, registration

Received October 10, 1996; revised January 20, 1997; accepted January 23, 1997

1. INTRODUCTION

A variety of physiological imaging tests have been developed and implemented clinically, including PET and SPECT

*Corresponding author

(e-mail: chuck.meyer@umich.edu;
 (web: <http://www.med.umich.edu:dipl>)

imaging. While these tests sometimes provide unique clinical information, their interpretation can be difficult without careful correlation and, ideally, registration with anatomical data sets. In addition, registration of test-retest physiological imaging studies is valuable for assessing regional physiological changes over time. We and others have used such fusion methods to assist in the diagnosis of hepatic hemangiomas

(CT and Tc RBC SPECT), lung cancer (PET and CT), brain tumor (fMRI and PET) and in the preoperative localization of foci of abnormal monoclonal antibody uptake (SPECT and CT), among others (Wahl *et al.*, 1993, 1994). While most would agree that image fusion is desirable and generally useful clinically, the maximal clinical value of multimodality imaging, e.g. the additional information obtained by combining the sensitivity and specificity of functional imaging with the structural resolution of anatomical imaging, can be fully realized only when accurate, multimodality registration becomes widely available at reasonable expense. Until then visual comparison of unregistered images will provide only marginal additional value. An even more valuable application of automatic registration may be the presentation to the diagnostician of previous comparison studies recomputed to match the current study's slice locations. In many cases the previous study represents a slightly different set of acquisition circumstances (contrast materials or choice of acquisition parameters) which may significantly alter the appearance of the two studies. An ideal registration algorithm should easily handle iso- or multimodality data set pairings.

A recent review article described the state of medical image and volume registration *circa* 1993 (van den Elsen *et al.*, 1993). The more useful and popular algorithms were based on surface matching (Pelizzari *et al.*, 1987, 1989) or minimization of variance in one data set within corresponding segmented regions defined in the registered data set (Woods *et al.*, 1993). Multimodality registration has evolved significantly in the last few years beyond earlier stages which required either user-biased homologous feature selection (e.g. the user-selected perceived common points on the two different studies) or tedious data preprocessing such as segmentation for surface definition or tissue discrimination and stripping of outer cranial tissue layers. Within the last few years papers describing 'similarity metrics' have began to appear (Hill *et al.*, 1994). Such descriptors are statistical in nature, as is mutual information. Collignon described the value of joint entropy as a quantitative registration metric (Collignon *et al.*, 1995b) and perceived it to be a generalization of Woods' earlier work (Woods, 1993). Viola and Wells described the value of the mutual information for multimodality registration only months later (Viola and Wells, 1995) as applied to computer vision topics. Within months other authors also described mutual information as a criterion for medical data set registration (Collignon *et al.*, 1995a; Studholme *et al.*, 1995; Wells *et al.*, 1995). A more recent journal publication describes multimodal volume registration and surgical applications (Wells *et al.*, 1996). Papers examining the resulting accuracy and capture ranges obtained with mutual information-based registration methods have just started to appear (Studholme *et al.*, 1996). Many more will undoubtedly appear soon.

Our purpose here is to briefly describe our implementation of the algorithm and demonstrate the typical ease, accuracy and computation times associated with the use of this algorithm on a moderately high-performance workstation. We have evaluated the algorithm in phantom studies as well as in selected cases where few other algorithms could perform as well. These registrations are meant to demonstrate the value of this new method without comparison of its accuracy relative to other existing techniques. Accuracy and appropriate case usage issues can best be evaluated in clinical trials that compare algorithm results using diverse clinical and modality conditions across multiple institutions.

2. METHODS

The entropy of a data set is defined as its average information content, while joint entropy is the average information of two data sets. The joint entropy, $H(a, b)$ of two data sets, a and b , is related to mutual information of a and b , $I(a, b)$, by the following classical relationship:

$$H(a, b) = H(a) + H(b) - I(a, b)$$

where $H(a)$ and $H(b)$ are the individual entropies of data sets a and b respectively (Papoulis, 1984). As can be seen in the classical relationship, mutual information, $I(a, b)$, is the amount by which the sum of the individual data set entropies must be reduced to account for correlations that exist between the two individual data sets (I is never negative). The previous expression defining mutual information in terms of entropies reduces to the following equation for mutual information in terms of the more fundamental probability density functions for data sets a and b ,

$$I = \sum \sum p(a, b) \log_2(p(a, b) / p(a)p(b)).$$

In implementation a geometrically transformed version of one of the data sets, called the homologous data set, is interpolated into the spatial frame of the other, called the reference data set. If the initial homologous data set is similar to the reference set, then there is little difference in performance between iteratively minimizing $H(a, b)$ or maximizing $I(a, b)$, since $H(a)$ and $H(b)$ are both fully visible in the reference frame and thus are independent of the registration; either method increases the correlation between the two data sets. However, if the initial pose is sufficiently different, such that the homologous data set is truncated in its transformation into the reference space, its entropy as measured in the reference frame is no longer constant with respect to registration variables. Under these conditions maximizing $I(a, b)$ results in a larger capture range for the registration process, because maximizing $I(a, b)$ encourages the homologous data sets to appear fully in

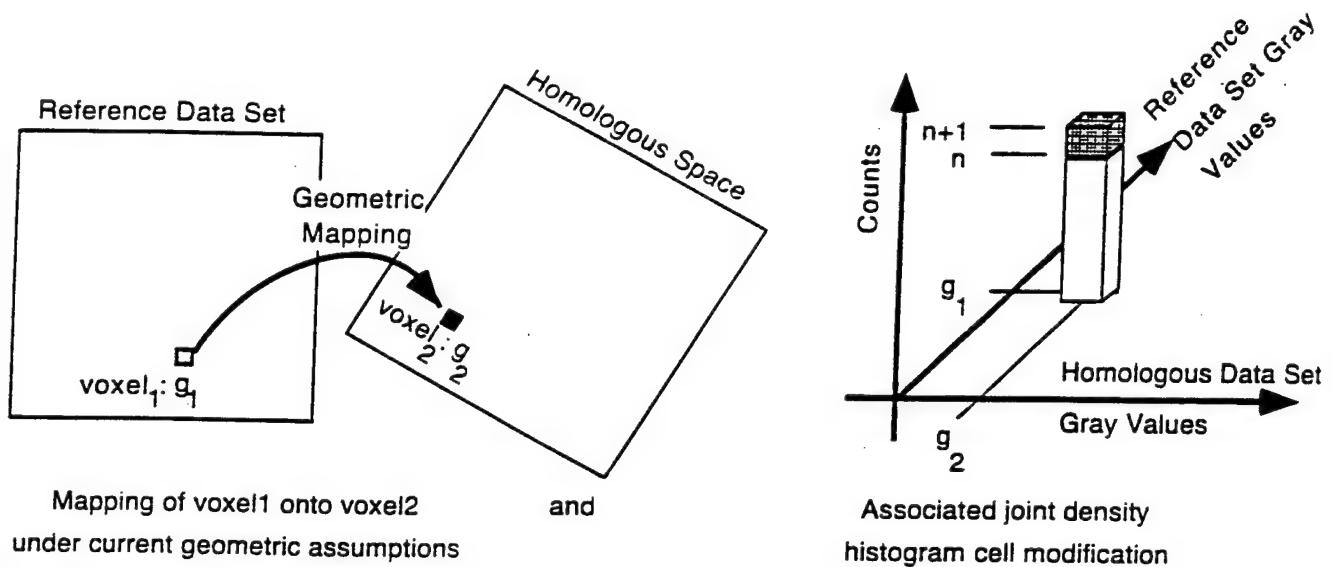


Figure 1. Method of constructing a joint density histogram for the two data sets.

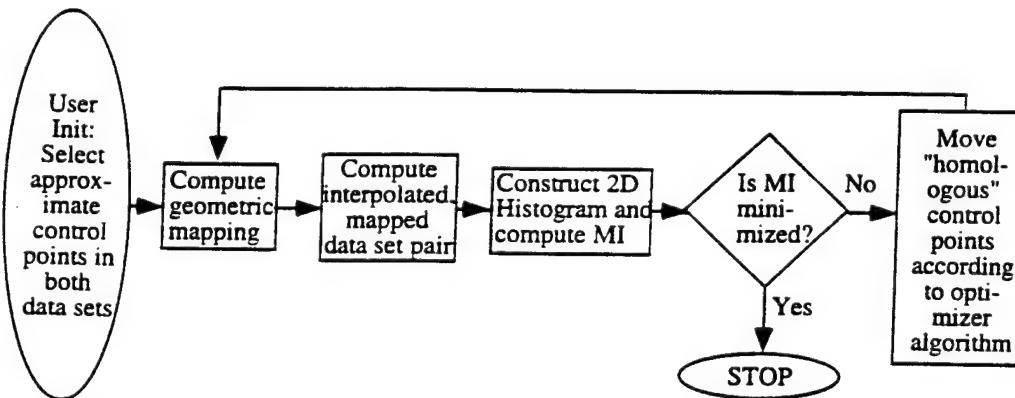


Figure 2. Functional flowchart of the automatic, MI-based, registration algorithm.

the reference frame due to its increasing contribution of $H(a)$ or $H(b)$ as it moves into the reference frame (anonymous reviewer).

Since we desire to use existing optimizer algorithms that traditionally minimize cost functions to obtain the condition of maximized I , we define our mutual information cost function, MI, to be related to I by $MI = -I$. If we choose the correct registration parameters and the two data sets are highly correlated, then I approaches its maximum and MI approaches its minimum; if they are uncorrelated, then MI approaches its upper bound of zero. Thus through registration maneuvers we can iteratively compute and minimize our cost function, MI. As depicted in Figure 1 our mutual information cost function, MI, is calculated directly from the joint

probability histogram of gray-values under a given registration's geometry mapping using the previous equation to form $-I$. The 2-D joint histogram is constructed by raster scanning through all voxels in the reference image set and incrementing bin counts corresponding to the gray-scale values from geometrically mapped voxel pairs. All registrations demonstrated here were achieved by a registration algorithm that uses the minimization loop shown in Figure 2. Note that the MI cost metric is computed using gray values from the reconstructed registered data set so that the effects of both the geometric mapping and pixel/voxel gray-value interpolants at all voxel locations are included.

The components of the control vector in our registration system are the 3-D coordinates of homologous control points:

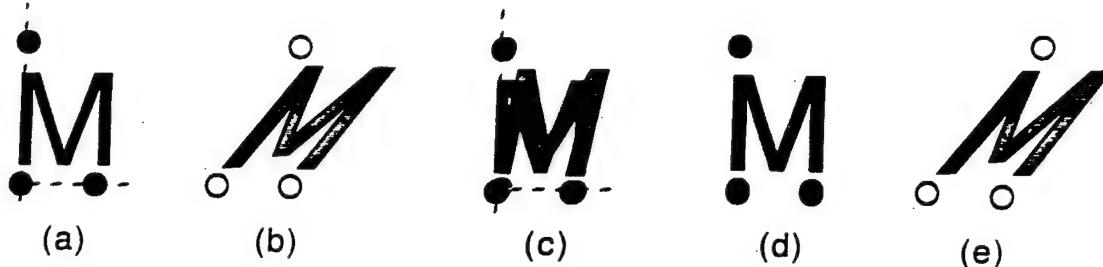


Figure 3. 2-D demonstration of intermediate steps in the registration algorithm including user initialization and algorithm optimization.

the resultant geometric mapping, i.e. registration, is computed from their positions. The interpolated gray-scale reconstruction is computed to match the reference data set, voxel for voxel, where the position of each voxel in the reference data set is mapped into the homologous data set using the previously computed geometric mapping. The gray value of each mapped voxel in the homologous data set is computed by trilinear interpolation of its eight nearest neighbors. Note that the interpolation process always uses the original data set to compute a new, geometrically mapped, gray-scale image for each iteration, instead of using the previous interpolation to generate the next one. This approach prevents round-off and other undesired interpolation effects such as smoothing from accumulating across iterations. When the geometric transformation maps to a position outside the field of view of the homologous data set, the resulting 'interpolated' voxel is set to the constant value of zero.

Affine, i.e. linear, geometric mappings are computed using four point pairs for 3-D data sets. In addition affine transforms can be constrained to varying degrees of freedom by decomposing the computed full affine matrix into the geometrically closest (Nobel and Daniel, 1988) desired approximation, i.e. rotate-translate, rotate-isotropic-scale-translate and rotate-anisotropic-scale-translate. Thin-plate spline (TPS) geometric mappings, i.e. non-linear warpings, are computed using at least one more point pair than used for the affine mapping (Bookstein, 1991; Bookstein and Green, 1993).

Although the user selects which input data set will be the reference and chooses the initial set of control point pairs for the reference and homologous data sets as in previous feature-based (user-biased) TPS methods, the algorithm automatically refines the position of the control points in the homologous data set under the direction of the optimizer by moving them to minimize the MI cost function.

The multivariate minimization algorithm we chose to drive the minimization of MI is the Nelder-Mead simplex method, usually referred to as 'amoeba' (Press *et al.*, 1988). Since this algorithm is fully explained in many numerical analysis

texts, and is widely used, we forgo further explanation of its function and refer the reader to Press *et al.* (1988) for further details. While certainly not optimal in terms of number of iterations to descend to the minimum, the simplex method does not require estimates of derivatives, is quite robust and thus its use is appropriate as an investigative tool.

Figure 3 illustrates a simplified 2-D example of the registration process described in the preceding paragraphs. Figures 3a and b demonstrate two 'multimodal' images in which the user has poorly chosen approximate homologous points. Figure 3c shows the geometric registration that maps the corresponding markers onto each other after the user's initialization, but before automatic optimization begins. After initialization, although the markers are perfectly aligned, the images are still misregistered. Figure 3d shows the resulting final registration obtained by the optimizer driving the locations of the homologous points in Figure 3b to minimize the cost function, MI. Note that the position of the upper marker in Figure 3b has been moved by the optimization process to a new position as depicted in Figure 3e which yields the correct registration by minimizing MI.

Besides performing initialization (quick placement of initial approximate homologous markers) the user can set the stopping criterion for the optimizer and the geometric registration model to be used (affine, TPS and associated degrees of freedom). Computation times depend primarily on the size of the reference data set since the interpolator (trilinear for voxels and bilinear for pixels) computes a matching gray value for each pixel/voxel of the reference, but also depends on the geometric mapping used to achieve registration. Computation time for the TPS mapping is approximately twice as long as the affine mapping for just four control points. In addition, computation times for the TPS algorithm are proportional to the number of control points used. All timings reported here are for runs on a DEC Model 3000/500x OSF/Alpha with a clock rate of 200 MHz. All applications run under the AVS visual programming environment (Advanced Visual Systems, Waltham, MA). While good programming practices

were followed in software module coding, almost no attention has been focused on coding for increased computational speed. Standard algorithms, e.g. singular-value decomposition of matrices, were implemented using code from *Numerical Recipes in C* (Press *et al.*, 1988).

3. RESULTS

3.1. Thoracic phantom study: PET/CT

A rigid, multicompartiment thorax phantom (Data Spectrum Corp., Chapel Hill, NC) was scanned on both a PET scanner and GE Genesis non-helical CT scanner. The Siemens 921 PET scanner is a 47-slice tomograph with an axial slice spacing of 3.375 mm (15 cm field of view), intrinsic in-plane resolution of ~6 mm FWHM and an axial resolution of 4.5–5.5 FWHM. The resulting volume matrix has $128 \times 128 \times 47$ voxels. CT data sets, acquired using 3 mm thick slices with no gaps or overlaps, were subsampled to 256×256 image matrices before utilization in the registration algorithm to reduce memory requirements. In addition to the presence of ^{18}F -labeled 2-deoxy-D-glucose (FDG) at various concentrations in phantom compartments to simulate a normal patient's thorax–abdomen scan, iodinated ionic X-ray contrast material was added to the liver cavity as well as lung and liver 'lesions' to increase their X-ray attenuation. Three 'lesions', two lung and one liver, had higher FDG concentrations simulating FDG-avid lesions; the liver 'lesion' also included a water-filled void simulating an FDG 'cold' lesion. The lung volumes were packed with small polystyrene spheres and water to mimic lung fields in CT and PET. A total of eight vitamin E capsules, 100 IU, were rigidly taped on the phantom, four around the phantom's superior circumference and four more around its inferior circumference, for use as fiducials in quantitative assessment of registration accuracy. Before PET scanning the capsules were injected with FDG, but in two of the eight total fiducials the isotope leaked from the capsules so as to render them invisible or indistinguishable from artifact streaks on the PET reconstruction.

The emission PET acquisition was 10 min in duration and collected 14.4×10^6 net true coincidences. Since the PET reconstruction contained end slices with changing modulation and noise, two slices at each end of the 47-slice data set were eliminated. The remaining PET data were linearly converted to byte data. After the isotope had decayed sufficiently, the phantom was scanned with CT using 3 mm thick slices and with no skip between slices. The phantom was intentionally rotated by $\sim 12^\circ$ about the anterior–posterior axis on the CT scanning table. The CT data were linearly converted to byte data over the range of -1024 to $+1800$ HU.

Initialization of the reconstruction algorithm did not use the external fiducial markers; in each modality four non-coplanar

homologous points were casually picked (by 'casual' we mean within 2–3 cm of the true homology), two near the apices of the 'lung' volumes and two at the level of the superior margin of the 'liver' volume. Due to the rigid construction of the phantom and the known calibration of the two scanners, a six DOF affine reconstruction (rotate–translate) was selected.

The optimizer stopped when movements of < 0.01 mm were requested in the x , y and z component locations of all control points. To avoid entrapment by local minima, restarts using a randomized simplex around the vector state of the previous stop are repeated until successive absolute changes in MI stopping values were < 0.0001 . For each restart the simplex is initialized to the previous stopping vector plus additional vectors randomly distributed within a user-chosen, city-block 'radius' of the previous stopping vector, which for all cases reported here was 10 mm. The term 'run' refers to an initial or repeated automatic optimization that ceases when the stopping conditions are met. The joint histogram was computed using 256×256 bins.

Although the PET data set was used as the reference geometry during the minimization of MI, the inverse of the final geometric transformation was used to compute a PET data set that matched the CT geometry and is displayed here in Figure 4 (part 4) using a combination of two contrast windows to allow simultaneous visualization of the low-level 'body' and fiducial concentrations and the much higher 'left-ventricle' and 'lesion' concentrations. Because the clusters in the joint density histogram with the largest membership have the largest weight affecting the registration, the resulting registration is essentially insensitive to the presence of the fiducial markers whose volume is very small compared with the full thorax. The resulting locations of the fiducial vitamin E tablets were measured by computing the centroid of the resulting fiducial intensity distribution. An error analysis for the six fiducials is shown in Table 1. The mean error vector length, i.e. the average of the square root of the sum of squares of Δx , Δy and Δz , was 3.84 mm and the combined standard deviation of all error vector components was 2.20 mm, resulting in a standard error of the estimated error vector of 1.56 mm. In comparison the PET voxel size is 4.385^2 by 3.375 mm³ with a mean voxel dimension of 4.05 mm. Thus these error measurements made just beyond the periphery of the PET data set (and thus beyond the region of support for the registration) are just slightly smaller than a voxel with essentially equal contributions from all components. The rotational errors are computed from the centroid of the phantom. The combined angular error mean and standard deviation for all fiducial components is -0.17° and 1.27° , respectively.

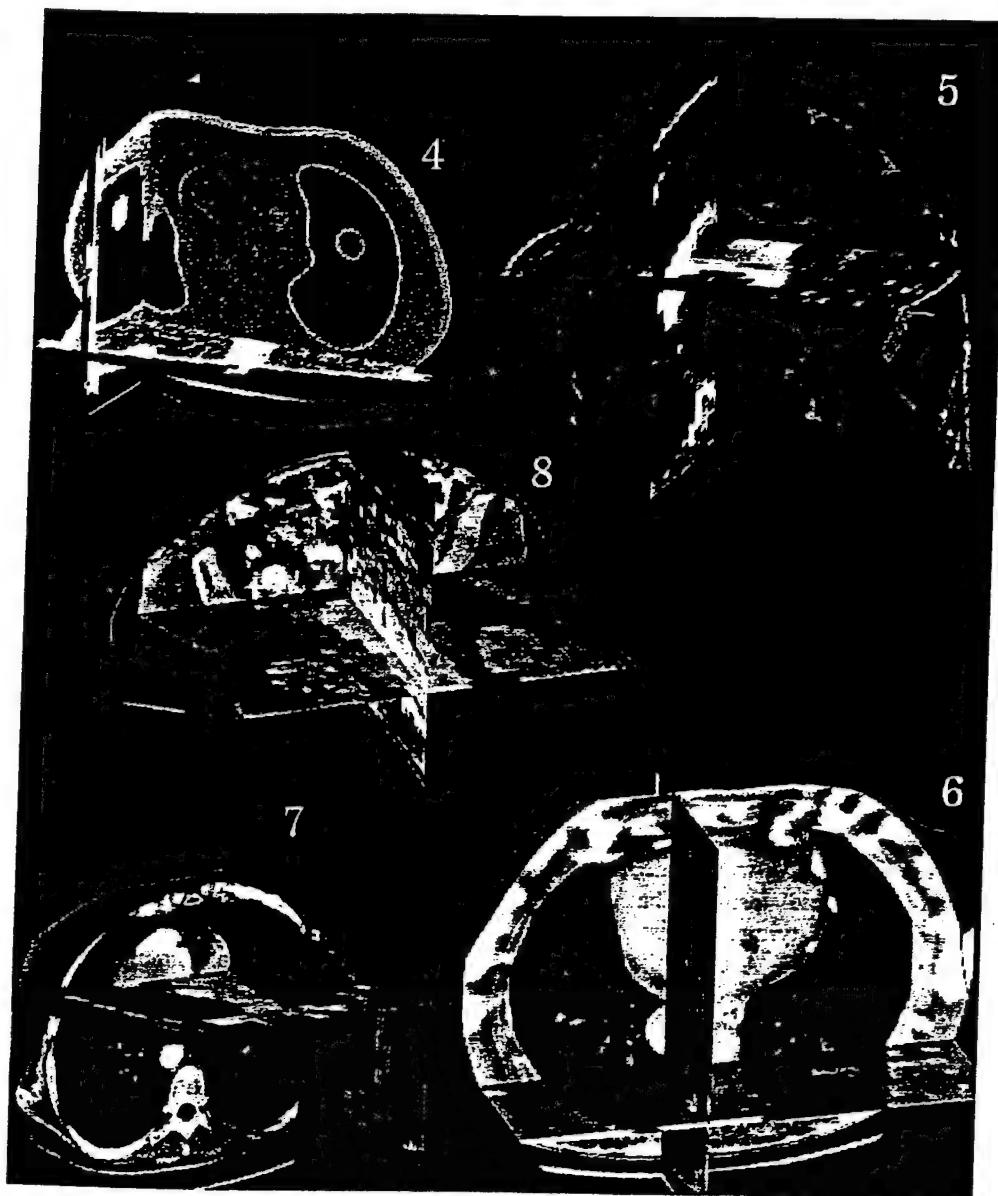


Figure 4. (Part 4) A 3-D rendering of three orthogonal slices through the thorax phantom's PET/CT registration is shown using CT as a reference to demonstrate partial-volume registration quality. The color composite combines CT data in gray and PET data in red. For PET a combined contrast window has been applied to demonstrate both low and high concentrations without associated blooming of the high concentrations. Alternating slices of the registration can be viewed on the associated CD-ROM movie. (Part 5) A 3-D rendering of the registration of a PET cerebral blood flow study and a post-surgical coronal MRI study is shown using MRI as a reference. The color composite combines MRI in gray and PET in pseudocolor. All slices of the registration can be viewed on the associated CD-ROM movie. (Part 6) A 3-D rendering of the full affine thorax registration is shown using the downsized (256 × 256) CT as a reference. In the color composite CT is shown in gray using a lung window contrast setting and PET is shown in pseudocolor. All slices of the registration may be viewed on the associated CD-ROM movie. (Part 7) A 3-D rendering of a PET/CT thin-plate spline-warped registration of a patient's thorax is shown using the downsized CT as a reference. In the color composite CT is displayed in gray using a combined lung and soft tissue window contrast setting and PET is shown in pseudocolor. All slices of the registration may be viewed on the associated CD-ROM movie. (Part 8) A 3-D rendering of a SPECT/CT, rotate-translate registration is shown using the downsized CT as a reference. In the color composite CT is displayed in gray using a soft tissue window contrast setting and PET is shown in pseudocolor. In addition to the expected concentrations in the liver, spleen and kidneys, note the uptake in the lymph nodes which reside near the descending aorta and the iliac arteries below the aorta's bifurcation. All slices of the registration may be viewed on the CD-ROM.

Table 1. Difference in fiducial locations on PET and registered CT of the thorax phantom for each fiducial. The associated error vector length is computed in column 5 and the associated rotational errors are computed in columns 6–8. The standard deviation for each of the above columns is computed in the bottom row.

Fiducial	Δx (mm)	Δy (mm)	Δz (mm)	Error vector			
				length (mm)	$\Delta\theta_z$ (°)	$\Delta\theta_x$ (°)	$\Delta\theta_y$ (°)
1	0.46	-1.22	-1.13	1.73	0.39	0.91	-0.24
2	-3.83	0.01	0.79	3.91	1.24	0.32	-1.37
3	-2.09	-3.08	0.30	3.73	0.82	-3.12	0.09
4	-3.78	-2.91	-2.67	5.47	1.70	-1.38	-1.03
5	-0.87	2.22	-2.32	3.33	0.63	-0.50	1.64
6	2.08	-2.11	3.90	4.90	-0.72	-1.20	-1.29
s, std dev.	2.36	2.02	2.43	1.31	0.83	1.42	1.15

Table 2. Statistics, mean (μ), population standard deviation (σ), two standard errors of the mean (2 SEM) and probability (P), from Student's T -test that the samples come from a zero mean population, all computed from nine repeated rotate–translate affine registrations of T2- and T1-weighted axial MRI data sets from NLM's Visible Male using randomized starting poses.

Statistic	Residual rotations (°)			Residual translations (mm)		
	$\Delta\theta_x$	$\Delta\theta_y$	$\Delta\theta_z$	Δx	Δy	Δz
μ	-0.0356	-0.0176	-0.0713	-0.1124	-0.2935	0.0429
σ	0.0094	0.0573	0.0142	0.0067	0.0144	0.0512
2 SEM	0.0063	0.0382	0.0095	0.0045	0.0096	0.0341
P	0.0000	0.3845	0.0000	0.0000	0.0000	0.0362

3.2. Post mortem study

The National Library of Medicine's Visible Male data was used as a source of MRI T1- and T2-weighted head acquisitions. Since the subject was motionless, we assume that the data sets are almost perfectly aligned as acquired. Data matrix sizes were $256 \times 256 \times 33$, where each voxel had dimensions of $1^2 \times 5 \text{ mm}^3$. In this registration accuracy test the T2-weighted data was used as the reference and T1-weighted data was used as the homologous set. During user initialization control points were chosen that resulted in the intentional initial misregistration of the two data sets. More specifically the initial misregistration in θ_x , θ_y and θ_z was -17.5° , 16.5° and 24.4° , respectively, and the initial (mis)translation in x , y and z was 16.4 , -14.4 and -3.4 mm, respectively. Because the object was rigid, a six DOF affine reconstruction (rotate–translate) was selected. The optimizer was set to stop when movements of <0.01 mm were requested for all component locations of all control points. Restarts using a randomized simplex around the vector state of the previous stop were repeated until successive absolute changes in MI stopping values were <0.0002 . Assuming the initial orientations of the

two data sets were identical, a perfect registration would yield final, residual rotation and translation values of zero. This test resulted in final residual rotations of -0.023° , -0.017° and -0.034° for θ_x , θ_y and θ_z , respectively, and translations of -0.12 , -0.32 and 0.046 mm for x , y and z , respectively. The process was repeated for eight additional randomly selected starting positions, resulting in the statistics of Table 2 for the mean (μ), population standard deviation (σ), two standard errors of the mean (2 SEM) and the probability (P), from Student's T -test that the samples come from a zero mean population. Thus all of the resulting registration parameters are statistically different from zero at the 0.05 level except $\Delta\theta_y$. Note however, that the means and standard deviations of all six parameters are very small; the largest translation is less than a third of a 1 mm voxel, $\bar{\Delta y} = 0.29 \pm 0.01$ mm, mean ± 2 SEM.

3.3. Clinical studies

For all of the clinical studies which follow the optimizer stopped when movements of <1 mm (subvoxel for the reference modalities of PET and SPECT) were requested in the

x, *y* and *z* component locations of all control points. To avoid entrapment by local minima, restarts using a randomized simplex around the vector state of the previous minimum are repeated until successive absolute changes in MI stopping values were <0.0002 . Again, the term 'run' refers to an initial or repeated automatic optimization that ceases when the stopping conditions are met. The joint histograms were all computed using 256×256 bins.

3.4. Brain PET/MRI and missing data

In the first case we demonstrate the registration of a PET data set from a patient who had a PET cerebral blood flow study in 1990 which involved transient carotid occlusion using a balloon catheter and a post-surgical MRI study in 1992. The PET study, obtained under the occluded carotid condition resulting in a PET signal from approximately half of the ipsilateral side of the head, was used as the reference scan for registering the post-surgical MRI. The registration was constrained to a six degrees of freedom (rotate-translate) geometric mapping. An older Siemens model ECAT 931/08-12 was used for PET data acquisition. The ECAT 931 is a 15-slice tomograph ($128 \times 128 \times 15$ voxels) with axial plane spacing of 6.75 mm (10 cm field of view), intrinsic in-plane resolution of ~ 6 mm FWHM and an axial resolution of 7–8 mm FWHM. The MRI for the same patient was obtained from a post-surgical imaging study obtained approximately two years later. The original MRI data set was a coronal, T2-weighted, spin-echo series consisting of 16 images of 5 mm width and 1 mm separation (TR/TE: 3000/90, FOV = 20, 256×256 matrix, NEX = 1). Before registration the MRI data set was low-pass filtered in the 'high-resolution' coronal planes with a Gaussian point spread function of 7 mm to approximate the partial volume contributions inherent in the PET acquisition. After casual user initialization of control points but before automatic registration, the MI was -0.9249 ; after two automatic runs consuming a total of 5.1 min of CPU time for 370 iterations, the final MI was minimized to -1.0292 . The final geometric mapping obtained with the low-pass filtered MRI data, was applied to the unfiltered MRI data and resulted in the registration displayed in Figure 4 (part 5). Note the lack of translation error (bias) in the resultant registration despite the missing data.

Few algorithms are capable of performing the registration demonstrated in Figure 4 (part 5). Note that the coronal head study begins anteriorly at the eyes, but stops before imaging the occipital cortex. Between the limits of both the half-head PET and coronal MRI, conservatively less than a third of the head volume is jointly available for driving the registration. Yet the resulting reconstruction shows no apparent bias. In addition, only low-pass filtering of the MRI data was required; no other preprocessing, e.g. stripping off scalp and other non-cortical structures by segmentation, was necessary.

By way of comparison, most distance mapping-minimization techniques would require not only scalp and bone removal before surface definition, but would also result in a biased registration where the increased distance measures caused by the missing PET data would drive the registration to place the existing data in the center of the MRI data set. Computationally intensive cross-correlation methods could obtain an unbiased result, but in this multimodal situation cross-correlation must be applied to edge-enhanced images since gray-scale mappings are inconsistent between PET and MRI.

In contrast with the previous description of MRI low-pass filtering for registration of partial data sets, in other experiments we have found that MI-based registration of normal PET/MRI data sets with a full region of support is trivial and requires no low-pass preprocessing.

3.5. Body

3.5.1. PET/CT

The following two cases demonstrate automatic volumetric registrations in the body using PET/CT data sets. The PET acquisitions are acquired on a Siemens 921 scanner using ^{18}FDG which traces glucose metabolism *in vivo*. The description of the tomograph can be found in the previous phantom registration description and methods of image acquisition are as previously outlined (Wahl *et al.*, 1994). The CT data sets, acquired using 1 cm thick slices with no gaps or overlaps, were subsampled to 256×256 image matrices before utilization in the registration algorithm to reduce CPU memory requirements.

Consistent patient geometry

The first body registration case we present was acquired on the CT scanner with free, shallow respiration and arms down to mimic PET data acquisition conditions. The patient has a right lung lesion easily visible on both CT and PET. At initialization preceding automated registration, four starting control points in each data set were chosen: one approximately at the carina and three more at the level of the diaphragm: one on the anterior chest midline and one each at the posterior-left and posterior-right chest wall. The control point at the carina was intentionally misplaced ~ 2 cm laterally to the right of the carina in the CT data set while the carina was correctly chosen in the PET data set to demonstrate how initial control points may be inaccurately placed and still achieve registration. After user initialization and before automatic registration the MI was -0.3387 . The geometric registration chosen was a full, 12-parameter affine geometric mapping. Convergence was obtained in four runs requiring a total of 31 CPU minutes for 725 iterations yielding a final MI metric of -0.5289 . The resulting registration is shown in Figure 4 (part 6).

Although the PET data set was used as the reference during the minimization of MI, the inverse of the final geometric transformation was used to compute a PET data set that matched the CT geometry and is displayed here in Figure 4 (part 6). Note the accurate delineation of the cardiac and vascular structures as well as that of the lesion. Although our previous experience indicates that registrations defined by surface markers represent internal organ geometries poorly, one relatively well registered sternal surface marker can be seen in one of the slices. The apparent registration accuracy using the full affine model is a tribute to positioning the patient in a consistent scanning geometry for both modalities as well as the use of an accurate registration algorithm.

Inconsistent patient geometry

In the next study the CT data set was acquired with the patient's arms held up over the head and respiration held during scanning between breaths at maximum inspiration, while the PET data set was acquired under the usual arms down condition. Between these two acquisition conditions internal organs and chest wall structures have different complex deformations which cannot be modeled by an affine transformation.

The patient has a large lesion in the right chest wall that invades the lung space and is visualized on both CT and FDG PET. A five-point thin-plate spline (TPS) reconstruction was used to partially compensate for the non-affine geometric changes. At user initialization, control points were placed approximately at both lung apices, carina, anterior chest wall at the level of the carina and on the lesion. After user initialization the MI was -0.4874. After three automatic optimizer runs which consumed a total of 57 min of CPU time for 650 iterations, the MI was reduced to -0.5971. As before, although the PET data set was used as the reference geometry during the minimization of MI, the inverse of the final geometric transformation was used to compute a PET data set that matched the CT geometry and is displayed in Figure 4 (part 7). In addition to the lesion, the lung apices, chest wall and cardiovascular structures are well registered. Note the avid ^{18}FDG uptake seen both in the lesion and the left ventricle of the heart.

3.5.2. Abdominal SPECT/CT

The following case demonstrates registration of ^{131}I anti-CD-20 SPECT and CT. The SPECT scanner is a three-headed Picker prism that generates $64 \times 64 \times 38$ data sets where the slice thickness is 7.12 mm with no skip between slices. The SPECT data set was obtained from a single camera head scan of 360° using filtered backprojection. The patient was treated therapeutically with 100 mCi of ^{131}I labeled antibodies for non-Hodgkin's lymphoma 90 h prior to SPECT imaging (Kaminski *et al.*, 1993, 1996). The CT patient data was

acquired using the standard 1 cm slice thickness. The 512×512 CT matrix was subsampled to a 256×256 matrix to reduce memory requirements with little or no penalty due to the much lower voxel density SPECT data set. Initial control points in the approximate shape of an equilateral triangle with vertices located on the surface of the body wall were chosen at the level of the inferior tip of the liver. An additional fourth marker was placed near the socket of the left hip. The registration was constrained to be a rotate-translate mapping. After user initialization the MI was -0.8793. After two automatic runs MI was reduced to -0.9265. Due to the small size of the SPECT data matrix, total elapsed CPU time for the affine registration was only 3.5 min for 300 iterations. Again, although the final affine registration parameters were obtained using SPECT as the reference data set, the final registration is displayed in Figure 4 (part 8) using the CT data set as the reference.

Note that in addition to expected uptake in the spleen, kidneys and liver, there is increased uptake in the specifically targeted, enlarged abdominal lymph nodes that surround the descending aorta and iliac arteries in the abdomen and pelvis; there is little uptake in fat. External skin fiducials appropriate to each scanning modality were placed prospectively on the patient and can be partially seen in two of the slices. In general, the registration between the external fiducials is poor. Given the excellent agreement evident in other widely distributed structures, it is clear that the limited geometric model used, i.e. rotate-translate, is incapable of offsetting all the complex deformations between the two data sets. Fortunately for this application the registration algorithm is driven (weighted) by larger structures such that the most relevant internal organs of interest are well registered.

4. DISCUSSION

We have demonstrated and assessed the accuracy of a highly automated algorithm for 3-D image registration based on maximizing mutual information in both phantom and patient studies. Although not demonstrated here, full brain data sets such as PET and MRI are easily and rapidly registered without preliminary stripping of skull and skin. Even partial brain data sets are registered without bias by forming data sets with consistent partial volume effects using low-pass filtering and limiting registration to a rotate-translate geometry. Literally in the strict sense of the definition, it appears that as more information becomes available, restrictions can be relaxed and more DOFs can be supported. For registrations involving the comparatively low-entropy SPECT data set it was necessary to restrict the geometry model to the simplest, six DOF affine model. Attempts at allowing isotropic scaling (seven DOF) or

anisotropic scaling (nine DOF) produced poor results. On the other hand the higher entropy PET data support registration not only with a full, 12 DOF affine model, but also with a 15 DOF TPS warping model as well. Other published neuroscience 2-D applications involving autoradiography and light imaging routinely support 18 DOF TPS warpings (Kim *et al.*, 1997).

To the extent supported by the information content of the modalities involved, the geometry deformation model should be chosen to match the physical deformation between the data sets. In the inconsistent geometry example of PET acquired with patient arms down while respiring freely and CT acquired with arms held over the head while respiration was held, the full affine registration was a poor compromise. The optimizer moved the right lung apex marker too far cranially to try to position the lesion correctly. The compromise result (not shown here) was that neither the lesion nor the lung apices and shoulders were correctly registered. When the five-point TPS warping registration was computed, the global fit was much improved as can be appreciated in Figure 4 (part 7).

Total computation times are strongly affected by the size of the reference matrix, since the homologous data set is mapped onto each voxel in the reference data set. Thus an important time-saving strategy is to use the smallest data set (measured in numbers of voxels) as the reference to compute the geometry mapping parameters. After computing the last iteration, the higher resolution data set may be used as a reference and the low-resolution data set interpolated to match the high-resolution set by applying the inverse of the previously computed, geometric mapping. The increase in computation time for thin-plate spline warping over the full affine transform is due to the increased number of control points and use of the more computationally intensive TPS reconstruction algorithm. Even so, computation times using an MRI data set ($256 \times 256 \times 23$) as a reference and a thin-plate spline mapping of 11 control points (33 degrees of freedom) is only 15 s per iteration with convergence at 575 iterations requiring just under 2.5 h.

An important strategy to reduce further total registration computation time is to perform the initial registration runs using a decimated data set to bring the two data set geometries into close agreement, before using the complete reference data set to accommodate more subtle partial volume effects. Since the computational cost is related to the number of voxels in the reference data set, then decimating, i.e. downsizing, the reference data set by two in each of the three volumetric dimensions will result in run times shorter by nearly a factor of eight. This approach is particularly valuable when the initial poses of the two data sets are grossly different. Such an approach significantly reduces the total run times reported for the cases

previously reported in this paper. Currently we routinely perform two initial runs using a downsized reference, followed by as many runs as necessary with the full size reference to satisfy the stopping condition. Although we have previously shown that increasing the number of bins in the histogram increases the sensitivity of the MI cost function (Kim *et al.*, 1997), with downsizing it is important to reduce the number of histogram bins used in order to avoid obtaining sparsely sampled, unrepresentative distributions of the geometrically mapped data sets. For this paper we did not perform such downsizing maneuvers in order to simplify its presentation through the use of a single, simple algorithm with limited input parameter ranges.

The accuracy of the registration is affected by the signal-to-noise ratio (SNR) in the data sets. In the case of the CT/PET thorax phantom study presented in Figure 4 (part 4), the PET Poisson noise dominates. Although the accuracy of the resulting registration obtained at the periphery of the data set, mean error vector length of 3.84 mm, is slightly below the PET voxel size, other lower count, and thus more noisy, acquisitions would yield reduced accuracies, while count-rich acquisitions with less noise (volumetric acquisitions contain 6–8 times as many counts) would produce better accuracies than those reported here. Since affine registrations can be decomposed into translate, rotate, scale and shear components, rotational and scaling error components cause error vector lengths to increase at radial distances further from the centroid. Thus the errors measured just beyond the periphery of the PET data set for the phantom study are typically worst case; smaller errors are typically measured more centrally.

In line with the previous discussion of accuracy and SNR, the registration of two MRI data sets from the NLM's Visible Male results in much smaller error estimates than those obtained with the PET/CT phantom. Using the angular and translation error variances from Table 2 for the MRI T2/T1 registration, the estimate of the mean error vector length, which has a Maxwellian distribution (Papoulis, 1984), at a radius of 10 cm from the centroid is < 0.11 mm and < 0.05 mm at the centroid. Also note that the non-zero means may indeed be real, i.e. the original T1- and T2-weighted acquisitions do have a fractional voxel offset. The superior-to-inferior course of three 'strings' around the exterior surface of the head in the two original data sets provides x and y fiducials for the 1–4 pixel inspection and visual verification of fractional displacements at nearly all axial slice levels. The most noticeable displacement is in the AP direction corresponding to the y -axis direction of Table 2. These slight displacements may be the result of object motion between scans or imperfect gradient amplifier control between the two acquisition sequences.

In summary, the inherent benefits in using the mutual information cost function, MI include:

- The registration is simple and independent of operator bias. The algorithm uses the full dynamic range of the data without edge enhancement, segmentation, disarticulation or operator-selected morphologic features.
- MI is computed from volumetric gray-scale data, not just surfaces. This facilitates geometric mappings that include warpings which fit internal structures as well.

The recent work of two groups (Collignon, 1995a, b; Viola *et al.*, 1995; Wells *et al.*, 1995) has ushered in a new era in registration where both multi- and isomodality registrations are achieved automatically and easily by the same software algorithm with little fine tuning of control parameters. We have demonstrated several clinically interesting modality pairs registered with this robust, easily used algorithm, many of which have geometric inconsistencies between data sets that are difficult for other algorithms to register. This algorithm requires only minimal user initialization (no segmentation, surface definition or data stripping) and typically runs in minutes on currently available, high-performance workstations. Although more validation efforts are required, the apparent registration accuracy and robust convergence properties for all modality pairs demonstrated here along with ease of use suggest that algorithms of this class significantly increase the likelihood that multimodality registrations will be routinely available to aid clinical diagnosis in the near future.

ACKNOWLEDGEMENTS

The authors wish to thank an anonymous reviewer who provided very constructive comments which led to a significantly improved text. This work was in part supported by DHHS PHS NIH 1R01CA59412.

REFERENCES

Bookstein, F. L. (1991) *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge.

Bookstein, F. and Green, W. D. K. A. (1993) A feature space for edgels in images with landmarks. *J. Math. Imag. Vision*, 3, 231–261.

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P. and Marchal, G. (1995a) Automated multimodality image registration using information theory. In Viergever, M. A. (ed.), *Computational Imaging and Vision, Proc. 13th Int. Conf. IPMI*, Vol. 3, pp. 263–274. Kluwer Academic Publishers, Dordrecht.

Collignon, A., Vandermeulen, D., Suetens, P. and Marchal, G. (1995b) 3D multimodality medical image registration using feature space clustering. In Ayache, N. (ed.), *Proc. CVRMed'95*, Lecture Notes in Computer Science, Vol. 905, pp. 195–204. Springer-Verlag, Berlin.

Hill, D. L. G., Studholme, C. and Hawkes, D. J. (1994) Voxel similarity measures for automated image registration. In Robb, R. (ed.), *Visualization in Biomedical Computing*, SPIE 2359, 205–216.

Kaminski, M. S., Zasadny, K. R., Francis, I. R., Milik, A. W., Ross, C. W., Moon, P. C., Crawford, S. M., Burgess, J. M., Peery, N. A., Butchko, G. M., Glenn, S. D. and Wahl, R. L. (1993) Radioimmunotherapy of B-cell lymphoma with [131-I] anti-B1 (anti-CD20) antibody. *New England J. Med.*, 329, 459–465.

Kaminski, M. S., Zasadny, K. R., Francis, I. R., Fenner, M. C., Ross, C. W., Milik, A. W., Estes, J., Tuck, M., Regan, D., Fisher, S., Glenn, S. D. and Wahl, R. L. (1996) Iodine-131-Anti-B1 radioimmunotherapy for B-Cell lymphoma. *J. Clin. Oncology*, 14, 1974–1981.

Kim, B., Boes, J. L., Frey, K. A. and Meyer, C. R. (1997) Mutual information for automated unwarping of rat brain autoradiographs. *NeuroImage*, 5, 31–40.

Nobel, B. and Daniel, J. W. (1988) *Applied Linear Algebra*. Prentice-Hall, Englewood Cliffs, NJ.

Papoulis, A. (1984) *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York.

Pelizzari, C. A., Chen, G. T. Y., Halpern, H., Chen, C. T. and Cooper, M. D. (1987) Three dimensional correlation of PET, CT and MRI images. *J. Nucl. Med.*, 28, 683.

Pelizzari, C. A., Chen, G. T. Y., Speilberg, D. R., Weichselbaum, R. R. and Chen, C. T. (1989) Accurate three-dimensional registration of CT, PET, and/or MR images of the brain. *J. Comp. Assis. Tomogr.*, 13, 20–26.

Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1988) *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge.

Studholme, C., Hill, D. and Hawkes, D. (1995) Multiresolution voxel similarity measures for MR-PET registration. In Viergever, M. A. (ed.) *Proc. 13th Int. Conf. IPMI. Computational Imaging and Vision*, Vol. 3, pp. 287–298. Kluwer Academic Publishers, Dordrecht.

Studholme, C., Hill, D. L. G. and Hawkes, D. J. (1996) Automated 3D registration of MR and CT images of the head. *Med. Image Anal.*, 1, 163–175.

van den Elsen, P. A., Pol, M. J. D. and Viergever, M. A. (1993) Medical image matching—a review with classification. *IEEE Eng. Med. Biol.*, 12, 26–39.

Viola, P. and Wells, W. M. (1995) Alignment by maximization of mutual information. In *Proc. 5th Int. Conf. on Computer Vision*, IEEE 95CH35744, pp. 16–23.

Wahl, R. L., Quint, L. E., Cieslak, R. D., Aisen, A. M., Koeppe, R. A. and Meyer, C. R. (1993) 'Anatometabolic' tumor imaging: fusion of FDG PET with CT or MRI to localize foci of increased activity. *J. Nucl. Med.*, 34, 1190–1197.

Wahl, R. L., Quint, L. E., Greenough, R. L., Meyer, C. R., White, R. I. and Orringer, M. B. (1994) Staging of mediastinal non-small cell lung cancer with FDG PET, CT, and fusion images: preliminary prospective evaluation. *Radiology*, 191, 371–377.

Wells, W. M., Viola, P. and Kikinis, R. (1995) Multimodal volume registration by maximization of mutual information. In Kikinis, R. (ed.), *Proc. Medical Robotics and Computer Assisted Surgery II*, pp. 55-62. Wiley, New York.

Wells, W. M. I., Viola, P., Atsumi, H., Hakajima, S. and Kikinis, R. (1996) Multimodal volume registration by maximization of mutual information. *Med. Image Anal.*, 1, 35-51.

Woods, R. P., Mazziotta, J. C. and Cherry, S. R. (1993) MRI-PET registration with automated algorithm. *J. Comp. Assis. Tomogr.*, 17, 536-546.

3D Doppler Image Signal Quantification of Breast Masses

Paul L. Carson, J. Brian Fowlkes, Marilyn A. Roubidoux, Aaron P. Moskalik, Anurag Govil, Daniel Normolle[†], Gerald LeCarpentier, Shiney Nattakom, Jonathan M. Rubin

Department of Radiology and
+ Department of Biostatistics
University of Michigan Medical Center
Ann Arbor, MI, USA

Corresponding Author:

Paul Carson, Ph.D.
Professor of Radiology
Kresge III, Rm. R2315
Univ. of Mich. Med. Center
Ann Arbor, MI 48109-0553

Tel (313) 763-5884
Fax (313) 764-8451
EMail pcarson@umich.edu

Running Title: Quantitative 3D Doppler Imaging of Breast

Abstract

New measures of Doppler image signals were introduced and a preliminary study was performed to evaluate which of these and earlier measures would serve in the hypotheses for a similar, larger study of discrimination of breast cancer from benign masses. Twenty women with breast masses observed on mammography and going to surgical biopsy were studied. Eleven masses proved to be benign and 9 were malignant. Both 3D power mode and mean frequency Doppler were performed. To identify the mass and other regions of interest, vessels were displayed as rotatable 3D color volumes, superimposed on selectable gray scale/color flow slices. Doppler signals were recorded in each of 6 ellipsoidal regions of interest in and around the mass and 2 in normal tissues. Six measures were computed in each region, three from power mode, two from mean frequency and two from combinations of both. Radiologists rated the gray scale appearances of the masses on a scale of 1 to 5 (5 = most suspicious) for each of 6 conventional gray scale criteria. Of the individual vascularity measures in individual ROIs, the log speed-weighted pixel density and log power-weighted pixel density in the lesion internal periphery showed the greatest discrimination of malignancy although neither was statistically significant and not as good as the peak variables described below. The mean visual gray scale rating was the best discriminator overall, but two vascularity measures each made promising scatterplots in conjunction with the average visual gray scale rating. These vascularity measures were the log peak normalized power-weighted pixel density (peak NPD) and log peak mean Doppler frequency times the peak NPD ($v_M \cdot NPD_M$). Each of these values was the maximum in any one of the five ROIs most closely associated with the mass. A possible rationale for the relative success of these peak values is the blood signal normalization and the inhomogeneity of most breast cancers and the expectation that the highest velocities (shunting) and largest collections of blood are not necessarily in the same region in and around the tumor. Peak NPD of cancers varied with age, decreasing by a factor of 45 from 33 to 77 years.

Key Words: Blood, angiogenesis, blood volume, perfusion, ultrasound, Doppler, image processing, three-dimensional imaging, breast cancer, diagnostic imaging

Introduction

Visual assessment of ultrasound Doppler imaging (color flow imaging) may be aided by quantitative analysis of information provided in color flow images. Several quantitative and semiquantitative studies have been performed recently, e.g. (Cosgrove et al. 1993; Huber et al. 1994; Kedar et al. 1995; Carson et al. 1995; Meyerowitz et al. 1996) and others summarized by Meyerowitz et al.. These involved primarily mean frequency color Doppler imaging (f-CDI), including fraction of color pixels and mean Doppler frequency shift or mean calculated velocity. Using the mean color pixel value, Huber et al. achieved a much higher sensitivity and specificity for differentiation of carcinomas from benign lesions than they did with visual analysis. Meyerowitz et al. quantified information in the power mode color Doppler imaging (p-CDI). One of the indices they measured in regions of interest (ROIs) was the mean power of color pixels. Their other two indices and terminology, color pixel density and power-weighted pixel density, were the same as two of those we have been investigating (Carson et al. 1995).

To standardize breast lesion vascularity measures further, to evaluate information provided in the new power mode color flow images, and to aid visual analysis, we undertook a visual and digital analysis of information in a study including 3D ultrasound examinations of 20 breast patients scheduled for breast biopsy (Carson et al. 1997). Studies of these biopsy-proven 20 cases (4 fibroadenomas, 7 other benign masses and 9 cancer cases) were reviewed by radiologists using eight visual criteria of pulse echo characteristics and 8 visual criteria of power mode Doppler characteristics. 3D vascular/pulse-echo display provided a much stronger subjective appreciation of vascular morphology and allowed a somewhat better ultrasound discrimination of malignant masses than did the corresponding 2D images or videotapes (specificities of 85%, 79% and 71%, respectively, at a sensitivity of 90%). Only in the 3D display did the vascularity measures display a trend towards significance in this small study. The pulse-echo ratings were quite significant. That same visual pulse-echo analysis is employed

in this present report for comparison with the digital vascularity assessment. Much of the background of this study was discussed in the visual analysis paper (Carson et al. 1997).

Materials and Methods

General Methods

Data was collected from 39 subjects in 44 exams in this preliminary study. Subjects were selected after the decision to biopsy based on identification of a mass in mammography, clinical exams or simple gray scale ultrasound. The quantitative measures were calculated for only 20 different subjects due to the quality in some studies and the need to develop special color map conversion curves in others. Masses of very low suspicion, e.g., most fibroadenomas, were not included. Nine masses were malignant, four were fibroadenomas, and seven were other benign masses.

Each exam consisted of one or more power mode 3D scans and a frequency shift 3D scan of the mass and the surrounding tissue. The 3D scans were acquired using the system described in (Moskalik et al. 1995). This system consisted of a conventional ultrasound scanner (Spectra VST, Diasonics, Inc., Milpitas, CA), a modified mammography unit which applied mild compression to stabilize the breast and a motorized transducer positioner mounted to the mammography unit. A typical scan through an acoustically transparent window produced a 3D volume of ultrasound data of the suspicious region of the breast from a series of approximately coronal 2D images separated typically by a distance of approximately $3/f$ (mm), where f is the Doppler imaging frequency* in MHz. At the 6 MHz Doppler frequency of our 6-10 MHz linear array, the step size was 0.5 mm. Data from the 50-120 images per scan were then transferred to a workstation. With the linear array scanhead employed for this study, the images were typically 38 mm by 40 or 50 mm. Power mode Doppler ultrasound was performed at a PRF of 700 Hz, an ensemble length of 14, and the minimum sample length and maximum gain not producing frequent color noise. The same was done in color flow mode but with a higher PRF of 1000 Hz.

* The step size should, ideally, be approximately half the elevational 6 dB beam width.

At this setting there was some aliasing, which decreased the recorded mean velocity in some ROIs of some subjects. Signals produced in the two color flow modes were measured to quantitatively estimate vascularity features in various 3D regions of interest in and around the suspected masses.

Quantitative Measures

The Doppler signal power imaging mode (Rubin and Adler 1993; Rubin et al. 1994) and frequency shift imaging mode provide the data necessary to quantify the Doppler image signal characteristics listed below. These are not absolute measures of tumor flow characteristics, as they are dependent on a number of uncontrolled factors outlined in the discussion. The relation of these quantities to fractional blood volumes and perfusion are also discussed more extensively in (Carson et al. 1993; Rubin et al. 1994; Rubin et al. 1995; Rubin et al. 1997). These quantities were developed to aid in the object assessment of the role of Doppler vascularity imaging for a more definitive diagnosis of suspect carcinoma.

Quantities in Table 1 were calculated in each of the eight regions of interest (ROIs) defined in Table 2, where the borders of the masses were taken from the pulse echo image and palpation. N_b is the number of pixels with flow in them and N_t is the total number of pixels in the region of interest. P_i is the power Doppler value of the i th pixel and V_i is the Doppler speed value of the i th pixel. This measurement of speed is calculated from the mean frequency shift value assuming an isotropic distribution of vessels, because no directional information is obtained up to this time. To represent 100% blood for normalization of the blood signals, P_b is the average Doppler power value measured in a large vessel. NPD_i refers to a single region i , while NPD_M refers to the peak of NPD_i s among a defined set of regions. For $v_M \cdot NPD_M$, the maximum of \bar{v} over any region and maximum value of NPD over any region were taken for defined sets of regions.

3D displays were illustrated in the visual analysis report (Carson et al. 1997). The data were used to form 3D isosurface renderings of the vasculature which could be rotated and shown in correct positional relationship to gray scale slices of the 3D image. An example of the employed

display on a nonpalpable, 7 mm intraductal carcinoma is illustrated more fully on our Web page: http://www.med.umich.edu/ultrasound/breast_anim/develop/Case38.html.

A typical example of a 3D color flow image display without vascular isosurface delineation is shown in Fig. 1. The gray scale borders of this infiltrating ductal carcinoma are demarcated by a translucent ellipsoid representing Region 1 (defined in Table 2) for quantitative analysis. Software was developed to select an ROI from the 3D images as an ellipse or an ellipsoidal annulus whose size and position are selected by the user who is analyzing the images. This was done to create reproducible ROI measurements and to help standardize the method of data acquisition between different sets of data. The quantities were calculated for each case in each of the regions of interest in Table 2 that could be identified. The 3D image sets in both Doppler image modes were carefully compared to select, as well as possible, the same ROIs in mode p-CDI and f-CDI image sets.

Since the two sets of images were obtained from sequential series of 3D scans, spatial matching of p-CDI and f-CDI coded pixels was quite crude for the pixel-by-pixel calculation of SNPD. For the matching of the entire ROIs required for calculation of $v_M \cdot NPD_M$, the image set registration was generally adequate to compensate for relative motion of the tissue and hand held scanner frame. Normalization of the power signal was done with respect to the mean signal power centered in the largest vessel at or near the depth of the lesion.

Results

Although data has been collected for 39 subjects, 11 were presumed normal or the preliminary diagnoses were unconfirmed. Some early studies were not quantified correctly or the mass was missed on the 3D scans. Analyzable data for vascularity were obtained on 20 and 17, respectively (see Table 3 below). In all the malignant and fibroadenoma cases analyzed, a mass could be identified from gray scale ultrasound. Lesion detection was sometimes guided by the local vascularity. In a few of the benign cases, the mass could not be found ultrasonically and a region of relatively high vascularity at least 1.5 cm dia. was selected as the "lesion". Subjects in which the signals were analyzed quantitatively are given in Table 3.

As examples of some of the measured quantities, the ranges and means of the NPD and SNPD for one of the regions of interest, ROI 1, are shown in Table 4 for the three different lesion types. The signal was normalized to the signal from the largest, strongest signal area at about the same depth as the mass. ROI 1 is defined in Table 2 by an ellipsoidal annulus between the lesion boundary and 1 cm inside. There is considerable overlap for the individual results in this single ROI, but the means of NPD for malignant masses in this region are differentiated reasonably well from benign conditions, particularly those other than fibroadenoma. Means of SNPD from ROI 1 also are quite different for malignant than for both benign classifications.

The NPD and SNPD in Region 4 are not particularly good classifiers of cancer. They revealed the significant vascularity in the outer periphery of fibroadenomas. In ROI 4 that is 1 cm outside the mass, malignant masses are not well discriminated. Tables 4 and 5 reveal large ranges of the two measures, even in ROI 1, and suggest that single measures in individual ROIs are not good indicators of overall discrimination ability of ultrasound.

Stepwise linear discriminant analyses were used as an exploratory tool to identify likely variables and combinations of variables in specific regions of interest to discriminate between benign and malignant masses. The quality of a given discriminant function was evaluated by the crossvalidated misclassification rates and the importance of variables within those functions by a t-test on the parameter. The only variable consistently identified as a discriminator between benign and malignant masses was speed-weighted color pixel density (SWD) in Region 1, the internal periphery of the lesion. A tree-based classification method was also employed and it consistently selected only power-weighted pixel density (PD), also from Region 1.

More promising results were obtained when the highest value of a variable for any one of the mass-related regions was analyzed, rather than looking at a specific variable in a specified ROI. Two measures that made the most promising scatterplots in conjunction with the average of all 6 gray scale visual classifications were the NPD_M (peak Normalized Power-Weighted Pixel Density) in any of a subject's Regions 1,2,4,5&6 and $v_M \cdot NPD_M$, in any one of those same regions. Region 3 was the only other region associated with the mass. It was not included

because of its rare and inconsistent vascularity, although that should not have affected the peak values discussed here.

Fig. 2 shows such a scatterplot of NPD_M and the average of all 6 gray scale visual classifications for each subject. The malignant masses clustered near the upper right hand corner as expected. The benign mass on the upper right was an extremely rare bilateral diabetic mastopathy, that met many of our ultrasonic, mammographic and clinical criteria for cancer. The cancer with the lowest vascularity index had multiple small areas of flow in the proximal periphery of the mass, i.e., the vascularity appeared lobular for what turned out to be a lobular carcinoma.

Fig. 3 shows a scatter plot of $v_M \cdot NPD_M$ (peak v and peak NPD) and the average visual gray scale for each subject studied. Note that all the measures of malignant masses are clustered near the upper right hand corner and are fairly well differentiated from the benign measures. Discriminant analysis on cancer vs. all benign masses provided the classification lines $\log NPD_M = -1.55 GS + 2.4$, and $\log v_M \cdot NPD_M = -1.27 GS + 3.0$, where GS is the average gray scale rating. The sensitivities of 100% and specificities of 82% and 90%, for NPD_M and $v_M \cdot NPD_M$, respectively, are promising, but the statistical uncertainty in these numbers is large.

The vascularity of malignant lesions as measured by the log of maximum NPD_M among the five regions did decrease significantly as a function of age. The linear regression was $\log NPD_M = -4.47 - 0.028 \pm 0.007 \times \text{age (yr)}$. The intercept and slope of the fit were insignificant and significant, probability $< |t| = 0.22$ and 0.006, respectively. See Fig. 4. The decrease in NPD_M with age suggests that age might be a significant discriminator of cancer along with NPD_M and gray scale rating. However, the number of data points in this study is too small for meaningful 3-variable analysis. Other key ultrasound measures were not a significant function of age, e.g., NPD for benign cases, which had an insignificant but negative slope ($|t| > 0.12$). An increase in gray scale reading with age for benign cases was not significant nor was an increase in $v_M \cdot NPD_M$ with age for benign cases.

Discussion

It is important to note that the measures NPD_M and $v_M \cdot NPD_M$ do not represent true fractional blood volume or perfusion however, they do provide quantification of what the clinician sees as well as some normalization for machine settings and ultrasound attenuation. This may provide a more reproducible result than what is currently done clinically. We previously referred to NPD as the detected fractional blood volume B_V , and to $v \cdot NPD$ as the speed-weighted detected fractional blood volume, B_{vv} , or even perfusion (Carson et al. 1993). We are now using the acronyms NPD and $v_M \cdot NPD_M$, or their longer descriptive names, because of the potential for over-interpretation of the more vascularity-specific terms.

New measures of color Doppler image signals have been developed for potential assessment of tumor vascularity. The noise has been applied in 3D with recently developed display techniques to improve the sampling statistics and reliability of defining regions of interest in and around the tumor.

Speed-weighted pixel density (SWD) and power-weighted pixel density (PD) in Region 1 were identified as the best single discriminators in single ROIs. These observations are consistent with one recent criterion for discriminating malignant masses as those in which the vessels penetrate the mass (Bergonzi et al. 1993; Calliada et al. 1994; De Albertis et al. 1995). The discrimination by either of these single variables in Region 1 alone was poor. Those linear discriminant and tree-based classification analyses were hampered by requiring a measurement of a given classifier within a given region for all subjects. The observation that no measure in any one region was a consistent classifier in all the malignant masses might be expected, given the heterogeneity of cancer growth.

Anticipated tumor heterogeneity was the basis for considering only the peak of a given variable obtained from among five of the regions associated with the mass. The normalized measures NPD_M and $v_M \cdot NPD_M$, taken as the peak values of NPD and $v_M \cdot NPD_M$ separately over the five regions, provided the best discrimination on scatter plots as a function of gray scale

ratings. The relative success of $v_M \cdot NPD_M$ calculated in this manner is consistent with the possibility that, due to shunting and other flow phenomena, peak \bar{v} might be higher in one region associated with the tumor while high volumes of moving blood might be in another region. The specificity and sensitivity obtained using NPD_M or $v_M \cdot NPD_M$ in conjunction with mean gray scale rating were consistent with or slightly higher than those obtained with visual vascularity and gray scale ratings (Carson et al. 1997).

There is a problem of measuring \bar{v} without aliasing effects when it is desired to see as much of the vascular tree as possible in f-CDI as well as p-CDI modes. One could scan at higher PRFs to avoid aliasing, but then the number of vessels seen without contrast agents and the diversity of their orientations might be too small for the unknown Doppler angles to average out. An intermediate PRF was employed.

Decreased vascularity with age is to be expected in the normal breast, likely related to fatty replacement of glandular tissue in aging. The clear relation seen for NPD_M with age suggests that ultrasound vascularity measures as an indicator for cancer or for general tumor aggressiveness would be improved by normalizing for age.

While these results are promising, the wide ranges of the vascularity measurements indicate that some breast cancers have very little detectable blood flow due to either low volume or low velocities. Some malignant cases only had vascularity in Region 6, a region associated with the mass but more than 1 cm away, suggesting fine vessel branching upstream from the cancer or most active perfusion well outside the cancer.

A vertical line on either Fig. 2 or Fig. 3 would make a better discriminator than a horizontal line. This indicates that the combined gray scale score was a better discriminator of cancer than the single vascularity measures. Vascularity will not be the sole criterion in most breast cancers, but it should contribute to the diagnosis, as there was added discrimination provided by the vascular results. This additional discrimination was not statistically significant in this small set of subjects used as a training set for the multiple variables studied. These results do provide the basis for the hypotheses of a future study on an independent, test set of subjects.

The problem with absolute measures from most unnormalized power mode signal indices such as power-weighted pixel density (PD) and color pixel density in either mode, f-CDI or p-CDI, is that the absolute measures increase with system gain and output and decrease with depth in the tissues due to attenuation and the lack of sufficient signal level in color flow imaging to effectively utilize TGC. The power measures which are normalized to the power signal from the largest vessel at a representative depth range also probably can fluctuate strongly in some cases. While the current results are promising, it remains to be ascertained that the normalization will improve stability of the vascular measures. What is needed for the normalization to work well is that the normalizing vessel be larger than the color Doppler sample volumes and be at the right depth or depths. In this study, the normalization was usually based on a single measurement at a single depth closest to the region of interest. Work is progressing on efforts to define signals from 100% blood based on all the larger vessels in each given depth range (Rubin et al. 1997) and to eliminate signals from low-shear portions of vessels where abnormally high Doppler signals are produced by Rouleaux formation.

Acknowledgments

This work was supported in part by USPHS grant 1RO1CA55076 from the National Cancer Institute and by the U.S. Army Medical Research and Materiel Command under Contract No. DAMD17-96-C-6061. The views, opinions and/or findings contained herein are those of the authors and should not be construed as an official Department of the Army position, policy or decision. Mihra Taljanovic, M.D., Mark Helvie, M.D. and Dorit D. Adler, M.D. contributed significantly to the clinical examinations and interpretations. The authors are indebted to Nancy Thorson, R.T., for assistance with subject recruitment, records and procedures and to Claudia Koitch for assistance with manuscript preparation.

References

Bergonzi M, Calliada F, Corsi G, et al. Role of echo-color Doppler in the diagnosis of breast diseases. Personal experience. *Radiol Med (Torino)* 1993;85:120-3.

Calliada F, Raieli G, Sala G, et al. Doppler color-echo in the echographic evaluation of solid neoplasms of the breast: 5 years of experience. *Radiol Med (Torino)* 1994;87:28-35.

Carson PL, Govil A, Moskalik AP, et al. Lessons from digital vascularity measures in 3D power mode images of breast masses. In: Jellins J and Madjar H, eds. *Didactic Letures and Workshop Presentations, International Breast Ultrasound School (with Ninth International Congress on the Ultrasonic Examination of the Breast)*, Indianapolis IN, Sept. 28-Oct. 1, 1995: 186-191.

Carson PL, Li X, Pallister J, et al. Approximate quantification of detected fractional blood volume in the breast by 3D color flow and Doppler signal amplitude imaging. In: Levy M, McAvoy BR, eds. *1993 Ultrasonics Symposium Proceedings*, Piscataway, NJ: Inst. for Electrical and Electronics Engineers, IEEE Cat. No. 93CH3301-9, 1993:1023-1026.

Carson PL, Moskalik AP, Govil A, et al. 3D and 2D color flow display of breast masses, *Ultrasound Med. & Biol.* 1997 (in press).

Cosgrove DO, Kedar RP, Bamber JC et al. Breast disease: color Doppler US in differential diagnosis. *Radiol.* 1993;189:99-104.

De Albertis P, Oliveri M, Quadri P, et al. Retrospective analysis of color Doppler ultrasonography and flowmetry findings in solid nodular pathology of the breast. *Radiol Med (Torino)* 1995;89:28-35.

Huber S, Delorme S, Knopp MV, et al. Breast tumors: computer-assisted quantitative assessment with color Doppler US. *Radiology* 1994;192: 797-801.

Kedar RP, DO Cosgrove, IE Smith, JL Mansi, JC Bamber. Breast carcinoma: measurement of tumor response to primary medical therapy with color Doppler flow imaging. *Radiol* 1994;190:825-830.

Meyerowitz CB, et al. Quantification of tumor vascularity and flow with amplitude color Doppler sonography in an experimental model: preliminary results. *J Ultrasound Med* 1996;15:827-833.

Moskalik A, Carson PL, Meyer CR, et al. Registration of three-dimensional compound ultrasound scans of the breast for refraction and motion correction. *Ultrasound Med Biol* 1995;21:769-778.

Rubin JM, Adler RA, Fowlkes JB, et al. Fractional moving blood volume estimation using Doppler power imaging. *Radiology* 1995;197: 183-190.

Rubin JM, Adler RS. Power Doppler. *Diagnostic Imaging* 1993.

Rubin JM, Bude RO, Carson PL, Adler RS, Bree RL. Power Doppler: a potentially useful alternative to mean frequency-based color doppler US. *Radiology* 1994;190: 853-856.

Rubin JM, Bude RO, Fowlkes JB, et al. Cumulative power distribution function: technique for defining a stable intravascular point for normalizing fractional moving blood volume estimates with power doppler us. *Radiology*, 1997, accepted with revisions.

Tables

Abbr.	Full Name	Formula
PD	Power Weighted Pixel Density	$PD = \sum_{i=1}^{N_b} P_i / N_t$
NMPCP	Normalized Mean Power In Colored Pixels	$NMPCP = \sum_{i=1}^{N_b} \frac{P_i}{P_b} / N_b$
NPD	Normalized Power Weighted Pixel Density	$NPD = \sum_{i=1}^{N_b} \frac{P_i}{P_b} / N_t$
SWD	Speed Weighted Pixel Density	$SWD = \sum_{i=1}^{N_b} V_i / N_t$
\bar{V}	Mean Speed in Colored Pixels	$\bar{V} = \sum_{i=1}^{N_b} V_i / N_b$
SNPD	Speed and Power Weighted Normalized Pixel Density	$SNPD = \sum_{i=1}^{N_b} \frac{V_i P_i}{P_b} / N_t$
$v_M \cdot NPD_M$	Peak Mean Speed Times Peak NPD	$v_M \cdot NPD_M = \max(\bar{V}) \cdot \max(NPD)$

Table 1 Summary of studied measures from the power mode and frequency shift modes of the color flow Doppler images. (The four most promising quantities are underlined.) As can be seen, the first three measures in Table 1 are obtained only from the power mode images. The fourth and fifth are from the frequency-shift color flow images and the last two utilize data from both modes. Symbols are defined in the text.

1. Lesion periphery. Normally from the outer gray scale border to 1 cm inside
2. Lesion periphery, proximal side only (hemi-ellipsoidal version of region)
3. Central core of those lesions which are >2 cm in diameter
4. External periphery -- mass gray scale boundary to 1 cm outside
5. External periphery, proximal side only (hemi-ellipsoidal version of region 4)
6. Abnormal Region of Highest Vascularity (>1.5 cm diameter and > 1 cm from mass, but associated with it)
7. Normal fatty tissue volume
8. Normal fibroglandular ROI

Table 2. List of the eight regions of interest used in the quantitative analysis of the 3D data.

The measures listed in table 1 are calculated from each region. Regions 1-6 are associated with the mass. The other regions, 7 and 8, did not show early promise and were not analyzed extensively.

	Studied	Analyzed for NPD_M	Analyzed for $v_M \cdot NPD_M$
Presumed Normal	8		
Fibroadenoma	7	4	3
Benign	8	7	7
Malignant	13	9	7
Unconfirmed	3		

Table 3. Subjects examined and analyzed.

NPD (Lesion periphery to 1 cm inside mass)				
	N	Min	Max	Mean
Benign	8	0.00000	0.00086	0.00029
Adenoma	4	0.00001	0.01088	0.00324
Malignant	9	0.00000	0.03718	0.00577
SNPD (Lesion periphery to 1 cm inside mass)				
	N	Min	Max	Mean
Benign	7	0.00000	0.00193	0.00038
Adenoma	3	0.00000	0.00244	0.00097
Malignant	7	0.00000	0.03800	0.00618

Table 4. Examples of two mass quantities for ROI 1, the lesion internal periphery.

NPD (Lesion periphery to 1 cm outside mass)				
	N	Min	Max	Mean
Benign	8	0.00002	0.00668	0.00167
Adenoma	4	0.00007	0.01500	0.00472
Malignant	9	0.00000	0.00684	0.00210
SNPD (Lesion periphery to 1 cm outside mass)				
	N	Min	Max	Mean
Benign	7	0.00000	0.00311	0.00078
Adenoma	3	0.00003	0.02582	0.00925
Malignant	7	0.00000	0.02541	0.00400

Table 5. NPD and SNPD measures from ROI 4, which is defined as an ellipsoidal annulus between the lesion boundary and 1 cm outside.

Figure Captions

Fig. 1. 3-D display from a 49 year old subject with an infiltrating ductal carcinoma. Vascularity detected in f-CDI mode is displayed in one of the actual scan planes, a), and in three orthogonal planes, b) and c). In b) and c), the borders of ROI 1 and ROI 4, respectively, are shown as smooth, translucent, green, ellipsoidal surfaces. c) actually shows a region a little smaller than ROI 4. The large and small arrows in b) show, respectively, the line of contact of the scan head in the actual scanned plane, 4 cm long, and a small amount of vascularity in the scan plane which is encompassed in ROI 1.

Fig. 2 Scatterplot of peak NPD_M and gray scale readings. Symbols are: benign -- circles; fibroadenomas -- triangles; cancer--diamonds. The linear discriminator is shown as the solid line.

Fig. 3. The digital vascularity index shown here, $v_M \cdot NPD_M$ and gray scale rating are over the regions used in Fig. 2 and symbols are as in Fig. 2.

Fig. 4. Cancer vascularity as a function of age, showing a factor of 45 decrease in peak NPD among Regions 1,2,4,5, & 6 from age 33 to 77. There is some uncertainty as shown by the (dotted) 95% confidence limits of the (solid) linear regression line.

Figures

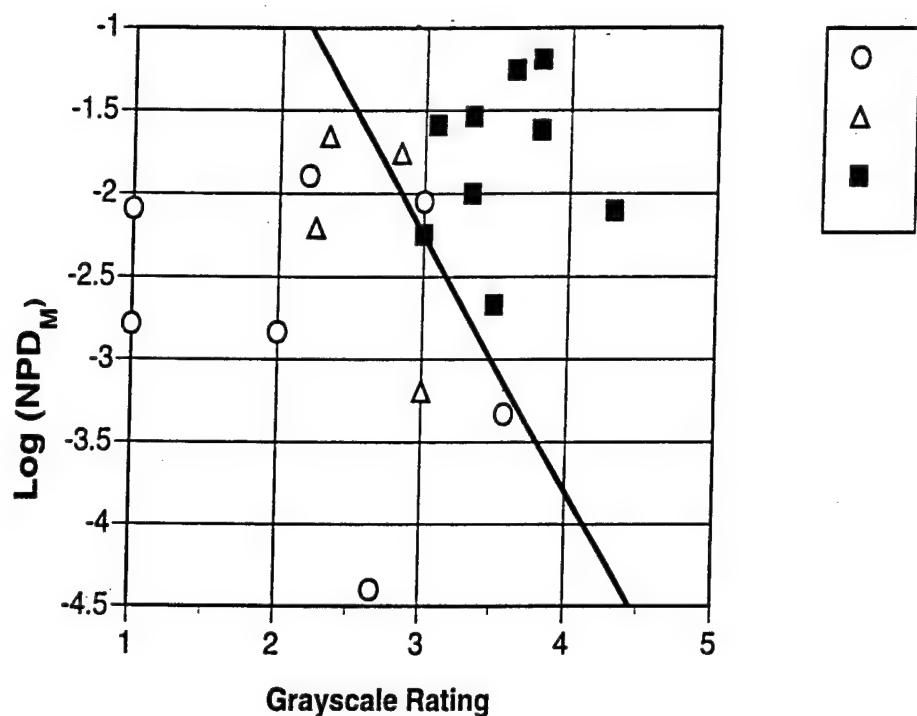
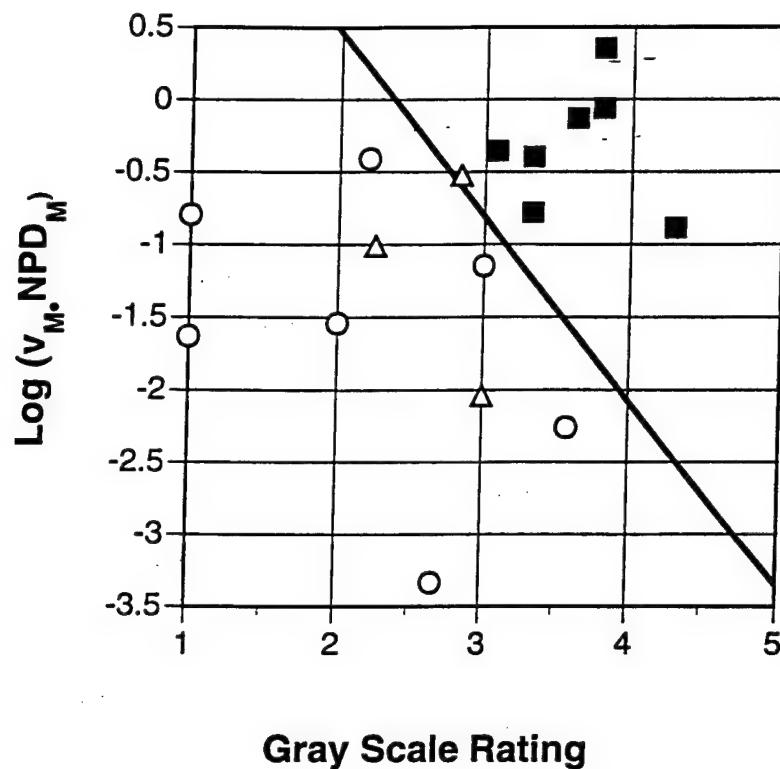


Fig. 2



Gray Scale Rating

Fig. 3

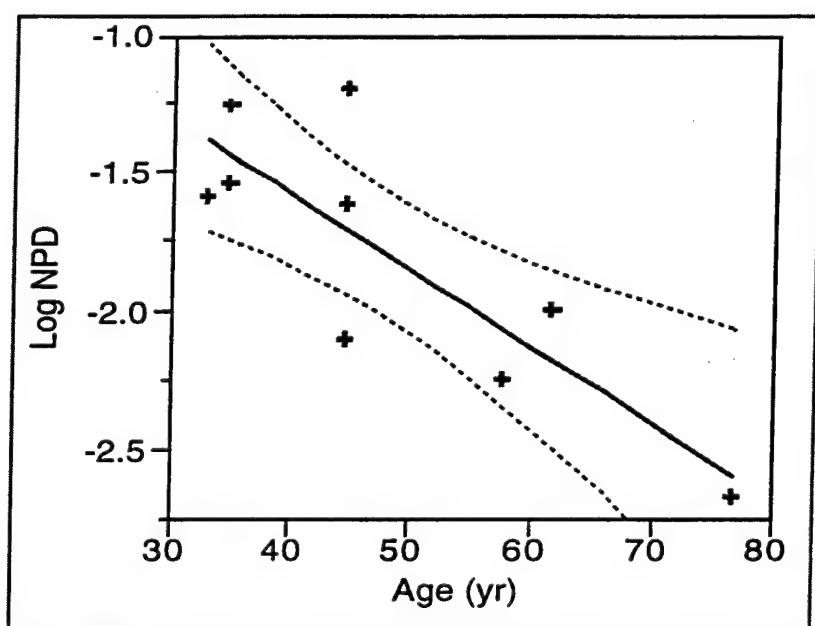


Fig. 4

3D Power Doppler Ultrasound Vascularity Findings in a Case of Breast Cancer

Shiney T. Nattakom, M.D.¹, Gerald L. LeCarpentier, Ph.D.¹,
Marilyn A. Roubidoux, M.D.², Paul L. Carson, Ph.D.¹

¹Department of Basic Radiological Sciences

²Department of Radiology

University of Michigan Medical Center

1500 East Medical Center Drive

Ann Arbor, MI 48109-0030

¹(313) 763-5692

²(313) 936-4555

Send all correspondence to:

Marilyn A. Roubidoux, M.D.

Assistant Professor

Department of Radiology

University of Michigan Medical Center

2910G Taubman Center

Ann Arbor, MI 48109-0030

(313) 936-4555

Fax: (313) 936-9723

3D Power Doppler Ultrasound-vascularity findings in a case of breast cancer

Shiney T. Nattakom, MD, Gerald L. LeCarpentier, Ph.D., Marilyn A. Roubidoux, MD,
Paul L. Carson Ph.D..

Although mammography remains the single most effective screening technique for early detection of breast cancer, sonography of breast lesions is becoming more common. In an attempt to more accurately predict whether a lesion is benign or malignant, sonographic features such as margins, echo texture and orientation to chest wall have been used¹. Vascularity of solid breast lesions has been studied with Doppler ultrasound²⁻⁴. There is an increased interest in the evaluation of vascular patterns in solid breast lesions with power Doppler ultrasound⁵. 2D and 3D power Doppler ultrasound findings in a young woman with breast cancer are presented.

CASE REPORT

A 42 year old woman diagnosed with bronchial carcinoid of the right lung underwent right middle lobectomy and bronchoplasty. Two of the five parabronchial lymph nodes were positive for tumor and she received radiation therapy. On a follow up chest CT a 1.8 cm lobulated soft tissue nodule was seen in the outer right breast (fig 1). She had no risk factors for breast cancer and had biennial mammography, with a mammogram 22 months previously being normal. A mammogram performed to evaluate the chest CT finding identified the new 2x2.5 cm lobulated mass in the upper outer quadrant of the right breast at the 10 o' clock position, without microcalcifications (Fig. 2). This mass was vaguely palpable.

A power mode ultrasound scan was performed with a GE Logiq 700 using a 7 MHz transducer. It showed a hypoechoic, mass with partially distinct margins and peripheral vascularity (Fig 3A). The device used to guide the manual scan and acquire the position information for 3D study is described in Fenn et al⁶. 3D reconstruction and isosurface rendering

of the vessels was done using the technique described in Moskalik et al⁷. Vessels inside the mass were more apparent in the 3D reconstructed arbitrary slice obtained almost parallel to the skin (Fig 3B & 3C). Isosurface rendering of vessels helps to demonstrate the continuity of vessels and their overall structure in relation to the mass (Fig 3D).

Subsequent wire localization and biopsy resulted in pathologic diagnosis of invasive ductal carcinoma (2.3 cm in maximum dimension) extending to the surgical margins with prominent angiolympathic permeation and positive estrogen and progesterone receptors. A modified radical mastectomy with reconstruction was performed.

DISCUSSION

Breast cancer is the leading cause of death in women in their forties. Mammography remains the most effective screening technique for breast cancer. As in many other screening techniques mammography too has a large "gray area" of indeterminate features. The positive biopsy rate for cancer is low, between 10% and 31% ⁸⁻¹⁰, i.e., 70%-90% of breast biopsies are performed in women with benign disease. Furthermore, in young women with dense breasts mammography is less sensitive than in older women with fatty breasts. The American Cancer Society now recommends annual mammography in the age group 40-49 rather than every 1-2 years. Critics contend that benefit in this population is difficult to translate to a clinical situation, in which compliance and radiological skill can vary. Mammographic surveillance rather than surgical biopsy is an acceptable method for follow up of mammographically "probably benign" lesions. Since all benign lesions cannot be confidently categorized by mammography some investigators have evaluated breast masses using gray scale, color Doppler and power Doppler.

Angiogenesis is necessary for rapid *in vivo* tumor growth. Power Doppler ultrasound measures amplitude of blood flow rather than velocity or frequency shift as in conventional color Doppler ¹¹⁻¹³. The major advantage of power Doppler is increased sensitivity to flow. This is due to relative lack of angle dependence, and the fact that noise has a low amplitude, thereby allowing the operator to increase the gain. The morphology and pattern of distribution of blood

vessels in solid breast masses seen at power Doppler ultrasound is a potentially important feature to be considered along with other sonographic criteria to predict the likelihood of malignancy⁵. Presence of irregularly branching and penetrating vessels has been reported to be correlated with malignancy^{5,14-16}.

However, study of the vascular architecture in 2D has its own limitations. On static 2D images, penetrating vessels lying near the exterior surface of the lesion normal to the plane of the image can be missed. Some vascular morphology, especially the continuity, direction and shape of the vessels can be much better appreciated with 3D imaging¹⁷. Fig 4 shows the different 3D reconstructed planes which are possible. An arbitrary slice may be reconstructed at any angle, whereas an arbitrary slice taken parallel to the skin will be difficult if not impossible to obtain in 2D alone. Ideally the 3D vascularity should be displayed in different colors to distinguish the vessels inside from the vessels outside the gray scale boundaries. This will greatly improve appreciation of vascular morphology and will clarify the vascularity within the gray scale defined mass.

Findings in this case show a distinctive advanced vascular pattern ultrasonically in a 42 year old woman within 22 months of normal mammogram. The usage of 2D and 3D scanning may improve the diagnosis in that age group. A study performed by Raza & Baum optimistically reported no false negative cases using a combination of gray scale & Power Doppler findings⁵. 2D scan plane orientation is influenced by physical constraints, operator preferences and experience. However fully reconstructed 3D views allow the selection of arbitrary image planes with unique and potentially important vascular information, unavailable from the 2D scan planes alone. In conclusion, gray scale & 3D Power Doppler could be a powerful combination in predicting the likelihood of malignancy in a solid breast mass.

ACKNOWLEDGMENTS

This work was supported in part by USPHS grant 1RO1CA55076 from the National Cancer Institute and by the U.S.Army Medical Research and Material Command under Contract No.

DAMD17-96-C-6061. J. Brian Fowlkes & Aaron P. Moskalik contributed significantly to the development of techniques.

References:

1. Stavros AT, Thickman DI, Rapp CL, et al: Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. *Radiol* 196:123-134, 1995
2. Dixon JM, Walsh J, Paterson D, et al: Colour Doppler ultrasonography studies of benign and malignant breast lesions. *Br J Surg* 79:259-60, 1992
3. Madjar H, Prompeler HJ, Sauerbrei W, et al: Color Doppler flow criteria of breast lesions. *Ultrasound Med Biol* 20:849-58, 1994
4. McNicholas MM, Mercer PM, Miller JC, et al: Color Doppler sonography in the evaluation of palpable breast masses. *American Journal of Roentgenology* 161:765-71, 1993
5. Raza S, Baum JK: Solid Breast Lesions: Evaluation with power Doppler US. *Radiology* 203:164-168, 1997
6. Fenn RC, Fowlkes JB, Moskalik AP, et al: A hand- controlled, 3-D ultrasound guide and measurement system. In S.Lees(Ed):*Acoustical Imaging*, vol.22. New York, Plenum Press, 1997, In press
7. Moskalik A, Carson PL, Meyer CR, et al: Registration of three-dimensional compound ultrasound scans of the breast for refraction and motion correction. *Ultrasound Med. Biol.* 21:769-778, 1995
8. Gisvold JJ, Martin J.K.Jr.,: Prebiopsy localization of nonpalpable breast lesions. *AJR* 143:477-481, 1984
9. Rosenberg AL, Schwartz GF, Feig SA, et al: Clinically Occult Breast Lesions: localization and significance. *Radiology* 1987:167-170, 1987
10. Basset LW, Liu TH, Giuliano AI, et al: The prevelance of carcinoma in palpable vs impalpable, mammographically detected lesions (comment). *AJR* 158:688-689, 1992
11. Bude RO, Rubin JM, Adler RS: Power versus conventional color doppler sonography: comparison in the depiction of normal internal vasculature. *Radiology* 192:777-780, 1994

12. Babcock DS, Patriquin H, LaFortune M, et al: Power doppler sonography: basic principles and clinical applications in children. *Pediatric Radiology* 26:109-115, 1996
13. Rubin JM, Bude RO, Carson PL, et al: Power Doppler: A potentially useful alternative to mean frequency-based color doppler US. *Radiology* 190:853-856, 1994
14. Bergonzi M, Calliada F, Corsi G, et al.: Role of echo-color Doppler in the diagnosis of breast diseases. Personal experience. *Radiol Med (Torino)* 85:120-3, 1993
15. Calliada F, Raieli G, Sala G, et al: Doppler color-echo in the echographic evaluation of solid neoplasms of the breast: 5 years of experience. *Radiol Med (Torino)* 87:28-35, 1994
16. De Albertis P, Oliveri M, Quadri P, et al: Retrospective analysis of color Doppler ultrasonography and flowmetry findings in solid nodular pathology of the breast. *Radiol Med (Torino)* 89:28-35, 1995
17. Carson PL, Moskalik AP, Govil A, et al: 3D and 2D color flow display of breast masses. *Ultrasound. Med. Biol. In press*

LEGENDS

FIG 1. Axial chest CT image demonstrates lobulated mass in the soft tissues of the right breast

FIG 2. Mammogram. Medial-lateral oblique view reveals a new lobulated mass (arrows) 2x2.5cm in dimension without microcalcifications.

FIG 3. 2D & 3D Power Doppler images of the mass. **A.** Hypoechoic mass with partially distinct margins and peripheral vessels demonstrated in a static 2D image. Bold white line on top represents the skin surface. **B.** 3D reconstructed image. Two planes shown here are scan plane (long arrows) and arbitrary plane (short arrows). **C.** 3D reconstructed image with scan plane (long arrows) and arbitrary plane (arrow heads) in a different position. The arbitrary plane clearly demonstrates vessels inside the mass. This plane is difficult to obtain in direct 2D scanning. **D.** Isosurface rendering of the vessels overlying the two planes of fig 3B shows vascularity inside the mass and continuity of vascularity.

FIG 4. 3D planes. Arrow 2 indicates the scan plane, arrow 3 the elevational plane (90° to scan direction), arrow 1, the plane parallel to skin (90° to the elevational plane and the scan plane), and arrow 4, a freely adjustable or an arbitrary plane.